AGRONOMIA LUSITANA

VOL. 16 — N.º 4

1954

ESTAÇÃO AGRONÓMICA NACIONAL

SACAVÉM

PORTUGAL

8 DEC 1955

AGRONOMIA LUSITANA

VOL. 16 — N.º 4

1954

Estação Agronómica Nacional

PORTUGAL

Composição e impressão das Oficinas
da Tip. Alcobacense, Lt. — Alcobaça

ESTUDOS DE ESTATÍSTICA AGRONÓMICA

V

OS MODERNOS INQUÉRITOS ESTATÍSTICOS POR AMOSTRAGEM NO ESTUDO DOS PROBLEMAS AGRÁRIOS NACIONAIS

POR *AUGUSTO JOSÉ DE OLIVEIRA*
(Estação Agronómica Nacional)

ÍNDICE

1 INTRODUÇÃO. OS INQUÉRITOS ESTATÍSTICOS POR AMOSTRAGEM 284
1a) Sobre o duplo papel do técnico estatístico no delineamento de ensaios e de inquéritos. Os Censos; os Inquéritos completos e por amostragem. O caso da «amostragem estratificada». 284
1b) Posição relativa dos ensaios experimentais e inquéritos por amostragem na investigação científica moderna 290
2 BREVE ESTUDO ESTATÍSTICO DOS RESULTADOS DO «INQUÉRITO ECONÓMICO E CULTURAL» REALIZADO EM 1930-32 À CULTURA DO TRIGO NA REGIÃO DO ALTO ALENTEJO 292
2a) Exposição sucinta de alguns métodos de tiragem de amostras e avaliação de estimativas 295
2b) Análise dos resultados do inquérito a explorações agrícolas do Alto Alentejo. O caso do «preço de custo do trigo» 312
3 OS ESTUDOS AGRÁRIOS POR AMOSTRAGEM EFECTUADOS OU A EFECTUAR EM PORTUGAL. DISCUSSÃO GERAL 322
3a) Indagações estatísticas levadas a efeito em Inglaterra, Estados Unidos da América do Norte e Índia . . . 323
3b) Outros pontos acerca dos inquéritos por amostragem ainda exemplificados com o caso do inquérito à cultura do trigo do Alto Alentejo 337
3c) Contribuição para o delineamento de inquéritos agrários por amostragem a realizar em Portugal e no Ultramar 340
4 SUMÁRIO. REFERÊNCIAS BIBLIOGRÁFICAS 344-345

1. INTRODUÇÃO. OS INQUÉRITOS AGRÍCOLAS POR AMOSTRAGEM (*)

1 a) ***Sobre o duplo papel do técnico estatístico no delineamento de ensaios e de inquéritos. Os Censos; os Inquéritos completos e por amostragem. O caso da «amostragem estratificada»***

COMPENETRADOS do interesse que pode revestir entre nós a aplicação das modernas técnicas da Estatística Matemática temos vindo a publicar alguns trabalhos estatísticos, uns de divulgação, outros com carácter prático, mas quase todos eles apenas respeitantes à *experimentação*.

Em boa verdade tal sector da pesquisa científica há mais de duas dezenas de anos que dispunha de técnicas próprias, filiadas numa doutrina lógica e estruturadas numa teoria matemática com base no cálculo das probabilidades; desde então, as Estações Experimentais Agrícolas e outros organismos de investigação vêm realizando os seus ensaios segundo delineamentos sugeridos por consultor estatístico dos respectivos quadros do pessoal, cuja intervenção passou cada vez mais a ser admitida como necessária, obrigatória.

Todavia, continuava a fazer-se sentir a falta de tratamento estatístico dos resultados de *inquéritos*, em particular dos de inquéritos por amostragem, apenas se dispondo quando muito de métodos empíricos ou arbitrários, até que alguns estatísticos se lançaram na remodelação das bases de amostragem e dos respectivos métodos práticos de tabulação e cálculo de estimativas.

Sob o imperativo da necessidade de resolução de muitos problemas práticos, dia a dia mais prementes, especialmente durante a 2.ª Guerra Mundial, e depois de inúmeras contribuições teóricas, chegou-se entretanto ao ano de 1946, data em que YATES anunciava à «Royal Statistical Society» a primeira síntese de tão importante matéria, a que tivémos já ocasião de fazer referência (OLIVEIRA, 1948).

Nesse trabalho focámos dum modo sumário os diversos métodos de amostragem, chamando especialmente a atenção para a denominada «amostragem estratificada» que tantas vantagens vinha já apresentando em relação à «amostragem casual» de outrora. A propósito dizia-se que o assunto estava sendo cada vez

(*) Um extracto da Introdução foi publicado no Boletim dos Estudantes do Instituto Superior de Agronomia («*Agros*», vol. 35, 1952).

mais aprofundado pelos estatísticos de todo o Mundo, quer nos seus aspectos teóricos, quer no campo prático, tendo até acabado de se formar na O. N. U. a «United Sub-Comission on Statistical Sampling».

O problema da amostragem nos inquéritos estatísticos fora no referido artigo apenas ventilado, tendo-se antes desenvolvido — através de exemplos numéricos com dados nacionais — os aspectos relacionados com a experimentação agronómica, resultando logo de tal conhecimento largos benefícios para a nossa Estação Agronómica Nacional e outros organismos agrários. *Já então se antevia o valor dos novos métodos de amostragem nos inquéritos, para cabal e rápido estudo dos problemas agrários*, ansiando-se nessa data por ver compendiada a respectiva metodologia estatístico-matemática, a fim de melhor permitir o seu estudo e compreensão.

Decorridos alguns meses entrava no prelo, finalmente, o livro de YATES (1949 a) destinado a, conforme se diz na introdução: «to cover all the modern developments of sampling theory which are of importance in census and survey work, and to give an adequate discussion of the complexities that are encountered in their practical application».

Pela primeira vez a matéria da amostragem era compendiada, embora duma forma mais pròpriamente descritiva, fazendo-se largo uso de exemplos numéricos tomados, aliás, do sector agrário — pelo que o livro, neste particular, se reveste do maior interesse para nós.

Das suas 300 páginas, precisamente as primeiras 150 são de carácter explanatório, nelas se distribuindo a matéria de 5 equilibrados capítulos:

«The Place of Sampling in Census Work»
«Requirements of a Good Sample»
«The Structure of Various Types of Sample»
«Practical Problems Arising in the Planning of a Survey»
«Problems Arising in the Execution and Analysis of a Survey».

Qualquer um destes temas é tratado pelo autor duma forma incisiva e completa, convindo que sejam largamente meditados por todos os técnicos que se disponham a planear inquéritos de responsabilidade, particularmente nos países atrasados ou com serviços de estatística mal organizados.

De acordo com o que importava numa primeira compendiação

de matéria tão vasta — aliás com alguns pontos teóricos ainda nebulosos — o autor apresenta-nos logo de começo a re-definição de muitos dos conceitos estatísticos a usar nas questões de amostragem, como os de «população», «amostra», «unidades de amostra» e «erros de amostragem», nestes englobando os erros sistemáticos (agora denominados «errors due to bias»), e os erros casuais ou fortuítos.

A propósito daquela sub-categoria de erros, os erros sistemáticos — na eliminação ou, pelo menos, na localização dos quais deve apostar-se o verdadeiro investigador bem como o técnico estatístico, do mesmo passo revelando com segurança a sua ciência e a sua arte de pesquisar — o autor cita exemplos de casos em que eles podem ocorrer, quer na tiragem da amostra, quer durante a própria interpretação dos resultados, não deixando de indicar os métodos defeituosos que a tal podem conduzir.

O processo rigoroso e mais simples para obviar àqueles erros sistemáticos, uma vez definida a população estatística em estudo, consiste na tiragem completamente casual da amostra, para o que se deve recorrer ao lançamento de dados, ou ao uso de Tabelas de Números Casuais, ou a qualquer meio mecânico apropriado. Assim, as estimativas que podem obter-se a partir de tal amostra apenas são afectadas de erros casuais de amostragem.

Acerca destes erros casuais também são apresentados métodos que permitem reduzi-los, a fim de se obter a precisão desejada, tais como os de: «estratificação», uso de «informação suplementar» e «amostragem por etapas múltiplas», cada um deles com a maior importância (particularmente o último, por ser mais prático), nos inquéritos por amostragem.

O capítulo 3, conforme se infere do seu título, trata mais pormenorizadamente da descrição da estrutura dos diversos tipos de amostra — já enunciados no nosso anterior trabalho — e dos métodos que devem ser seguidos na sua tiragem, a fim de garantir sempre a necessária representatividade. Tal matéria virá a merecer-nos maior desenvolvimento no capítulo seguinte deste artigo.

Nos capítulos 4 e 5 YATES refere as diversas questões práticas que se podem levantar no delineamento duma pesquisa por amostragem, que «vary greatly with the type of material and the nature of the information that is required» (*op. cit.*, pág. 48). Especial atenção é dada aos aspectos e circunstâncias que condicionam a definição do «universo estatístico» («the frame», na nova termi-

nologia de YATES), conforme o sector a inquirir e principalmente a variabilidade nele manifestada. Por fim apontam-se os delicados problemas respeitantes à organização administrativa de cada inquérito por amostragem, quase todos comuns, aliás, aos casos dos censos ([1]) e inquéritos completos.

As restantes 150 páginas do manual de amostragem que vimos referindo são preenchidas por outros 3 importantes capítulos, nos quais se tratam as matérias de valor mais pròpriamente matemático-estatístico:

«Estimation of the Population Values»
«Estimation of the Sampling Error»
«Efficiency»

com indicação das fórmulas a usar em cada tipo de amostra e método de avaliação de estimativas, seguindo-se a notação aconselhada pela «Sub-Commission on Statistical Sampling to the Statistical Commission of the U. N. O.» no seu *memorandum* da 2.ª Sessão de Geneva, de 1948 (UNITED NATIONS ORGANISATION, 1948 a) ([2]).

Também teremos ocasião de fazer largo emprego noutro capítulo de todas estas noções; desde já convém pôr em destaque que o termo «eficiência» corresponde ao de «precisão», havendo, todavia, alguma diferença quando se trate do caso de populações que não sejam grandes.

Pelos métodos apropriados (citados no capítulo 8 do livro de YATES) podem avaliar-se as eficiências relativas dos diversos tipos de amostragem, o que permite ao estatístico ajuizar dos méritos de cada um deles, e tanto mais completamente quanto maior for o número de dados referentes ao custo de cada uma das fases do inquérito.

Por fim citam-se cerca de 4 centenas de referências bibliográficas, distribuídas por 10 secções que abrangem os mais variados campos da investigação e até os dos negócios — onde o especialista poderá orientar-se devidamente.

Não se quer deixar de referir a advertência do autor, feita logo no prefácio, de que o livro teve de ser escrito um tanto à pressa e sob o imperativo da necessidade, por dele haver urgência

([1]) O termo inquérito refere-se mormente a indagações profundas ou de pormenor.

([2]) Um extracto do «Report» relativo à 1.ª Sessão de 1947 foi publicado em 1948, em Portugal, pela *«Revista de Economia»*, no seu Volume 1.º (Fasc. III).

antes de se iniciar o Censo Mundial de 1950, conforme os projectos da O. N. U., visto este organismo internacional ter chegado à conclusão de que as técnicas recentes de amostragem teriam de ser usadas, pelo menos nos países ou regiões mais atrasadas do globo.

Assim se justifica, em grande parte — em livro de tanta responsabilidade — a falta duma mais larga exemplificação destes métodos estatísticos não experimentais.

Em presença do referido manual: «*Sampling Methods for Censuses and Surveys*» mais fácil se torna agora a classificação geral dos métodos estatísticos experimentais e não experimentais. Cremos que deste modo se alarga imensamente o campo de investigação *e assim mais justamente deverá competir ao técnico estatístico o duplo papel de «delineador» de ensaios e de inquéritos*, conforme aliás um especialista eminente se pronunciara já em trabalho apresentado na 25.ª sessão do «Instituto Internacional de Estatística», em Washington (EISENHART, 1948).

A muitos pode parecer descabida a proposta do alargamento do uso da amostragem a âmbitos tão vastos como os abrangidos nos Censos e nos Inquéritos totais ou mesmo no Ordenamento — uma vez que na Experimentação Científica era já bastante usada, cada vez mais com predomínio das modalidades que implicam restrições à amostra do tipo casual, embora nem sempre com resultados totalmente satisfatórios, é certo, por via da multiplicidade e complexidade inerentes.

Só relativamente tarde se reconheceu que com as novas técnicas (algumas vezes apenas denominadas pelo nome genérico de «*técnicas de amostragem estratificada*») podia atingir-se esses domínios em que o emprego da simples amostragem casual se tornava quase sempre de difícil aplicação; e mais, que com elas muito se ganhava em tempo e em dinheiro — não deixando todavia de darem estimativas com a precisão requerida, ou seja, com um grau de confiança mensurável, facto este do máximo relevo (*Cf.* OLIVEIRA, 1948).

É que uma das maiores características da nova amostragem consiste na sua maleabilidade, bastando citar o caso da «amostragem de etapas múltiplas» (que pode incluir imensas variantes) pois, como muito bem refere YATES «has the important advantage that subdivision into second-stage units, *i. e.* the construction of the second frame need only be carried out for those first-stage

units which are actually included in the sample. It is therefore particularly valuable in surveys of undeveloped areas where no frame exists which is sufficiently detailed and accurate for subdivision of the material into reasonably small sampling units».

Em relação à Estatística oficial, Censos e Inquéritos totais os métodos de amostragem, em particular os mais modernos, apresentam ainda a vantagem de acréscimo de rigor — por maiores possibilidades de definição do «universo estatístico», indagações *in loco*, emprego de equipas especializadas ou mais cuidadosas, etc., permitindo ainda obviar aos inconvenientes da não-resposta, que tanto afecta aqueles (principalmense nos Inquéritos por correspondência, relativamente aos directos). Como principal desvantagem aponta-se, entretanto, a de não ser fácil usar contra-provas, defeito a que se pode obviar por meio de inspectores para isso adestrados.

Não queremos trazer aqui todos os casos da controvérsia que em dado momento se travou, pois isso nos parece já desnecessário, tanto mais que a eles se fez a devida referência, quer no nosso anterior trabalho sobre amostragem, quer num breve mas elucidativo artigo aparecido na *«Revista de Economia»* (GUERREIRO, 1948), no qual também se destaca o papel da amostragem e das suas imensas possibilidades, tornada mesmo necessária quando o universo estatístico a indagar for infinito ou hipotético.

De resto, essas técnicas modernas vêm sendo já usadas em Portugal no Instituto Nacional de Estatística (Ministério das Finanças) e no Subsecretariado de Agricultura (Ministério da Economia), devendo salientar-se o importante caso do emprego da amostragem no «Plano de Fomento Agrário» que, aliás, teve a colaboração do primeiro organismo e a que não fomos também estranhos na fase de planificação.

Embora pareça arrojada, cremos assim que continua a manter-se a afirmação, de certo modo contida no nosso trabalho de 1948, *de que por amostragem se podem conseguir resultados práticos e da maior precisão e rigor, ainda acrescendo a vantagem, quase exclusiva, de permitir a efectivação de grande número de Inquéritos à escala regional ou nacional, mesmo para se avaliarem as potencialidades do Património metropolitano ou ultramarino, todos eles estudos de grande folego que doutra maneira não seriam porventura possíveis nos tempos mais próximos.*

1 b) ***Posição relativa dos ensaios experimentais e inquéritos por amostragem na investigação científica moderna***

De passagem anote-se que no conhecimento da realidade não se deve sobrestimar o papel dos inquéritos em relação ao dos ensaios experimentais, antes haverá necessidade de torná-los complementares. Meditem-se estoutras afirmações produzidas na «United Nations Scientific Conference on the Conservation and Utilization of Resources» realizada em 17 de Agosto de 1949:

> «Surveys should be used to provide the basic background information, and to give partial information on the effects of factors that do not easily lend themselves to experiment, but factors which can reasonably be made the subject of experiment should be tested. *Only by experiment can results be obtained in which full confidence can be placed*» (o sublinhado é nosso).

Nelas o matemático YATES demarca claramente o valor relativo daquelas duas facetas da actividade de pesquisa científica; a única objecção que aqui podemos levantar diz respeito à natureza «política» das questões e problemas importantes — sujeitos como se acham, em última análise, a rígidas instituições jurídicas e económicas imperantes em cada momento histórico, e por esse facto cerceando ou mesmo impedindo a investigação de ir até ao âmago das coisas, para abrir os caminhos que verdadeiramente permitam modificar a realidade na consecução de melhores condições de vida geral. Aquele aspecto foi subestimado, a nosso ver, no trabalho apresentado pelo referido autor.

Por agora não pretendemos entrar mais a fundo neste palpitante tema das finalidades e possibilidades da Investigação Científica (1) mas fixe-se desde já o seguinte, porque apresenta o maior interesse para o caso agrário: a simples ideia de ensaios e inquéritos numa escala restrita como finalidade última e única, tem de ser, em grande parte, posta de lado. Dever-se-ão antes estabelecer esquemas estatísticos, quer experimentais, quer de inquérito, numa base ampla e geral (ultrapassando mesmo o sector limitado da produção ou os domínios meramente especulativos — como acon-

(1) Leia-se, entre outros, o livro de BERNAL (1943) e o artigo de YATES (1949 b) cujo título só por si é significativo: «Agriculture, Sampling and Operational Research».

tece ainda com a nossa Investigação); todos os factores de interesse prático, incluindo os que se liguem à lavoura, terão de ser tomados em conta.

Aqueles ensaios e inquéritos restritos (a que costuma dar-se o nome de «prévios» — os «pilot experiments» e «pilot surveys» dos ingleses) servem tão sòmente para permitir ajustamentos de pessoal e verbas e para dar a conhecer factores que virão depois a ser submetidos a estudo exaustivo. Então, nesta fase superior da pesquisa, os pequenos talhões de ensaios (no caso agrário) serão substituídos pelos campos da lavoura da região, em número elevado, e os inquéritos serão mais vastos e perfeitos, as questões mais aprofundadas.

Em todas essas fases o papel da amostragem continua a valorizar-se, conforme se depreende, pois por seu intermédio a investigação não redundará uma empresa muito dispendiosa. De resto, os grandes gastos em transportes e ajudas de custo, etc., com as equipas de múltiplos investigadores virão a ficar relativamente pequenos em face do alcance dos resultados que serão obtidos, ao mesmo tempo, gerais, práticos e seguros.

Por fim, o «operational research worker must be capable of integrating the evidence from all these sources into a coordinated whole, and presenting the relevant portions to the administrators (with whom there should be close and continuous collaboration) (YATES, 1949 b). E assim se irá passar com confiança ao domínio da «extensão» e do «fomento».

No presente trabalho procura fazer-se a análise estatística dos resultados do inquérito «económico e cultural» levado a efeito pela antiga Estação Agrária Central em 1930/32 (MINISTÉRIO DA AGRICULTURA, 1934) na Região do Alto Alentejo — únicos dados extensos até hoje publicados no campo agrícola, mesmo assim insuficientes (para o nosso fim em vista) e de restrito valor, por via da falta de representatividade da amostra em relação aos dois distritos do Alto Alentejo.

Primeiramente, antes da aplicação das técnicas estatísticas às diversas características das explorações agrárias inquiridas far-se-á uma exposição sucinta dos principais métodos de selecção da amostra e avaliação de estimativas (médias ou totais, e respectivos erros padrões) por cada tipo de amostragem, dando-se, também, as fórmulas de cálculo com exemplificações apropriadas.

Em seguida, virá a interpretação dos resultados do inquérito

sobre cada uma das variáveis estatísticas (*Cf.* Quadro A), principalmente as mais frequentes:

área de trigo da exploração agrícola (em ha)
despesa com a cultura de trigo (em escudos)
produção anual de trigo (em kg)
mão de obra assalariada certa (n.º de assalariados permanentes)
gado total da exploração (n.º de cabeças)

sendo analisado, por fim, o caso do «preço de custo de 1 kg de trigo» (a principal finalidade deste inquérito económico-agrícola), estratificando-se a amostra à base da grandeza da exploração agrícola, em primeiro lugar, e, depois, segundo a forma da exploração (Tabela III).

Considerar-se-ão, todavia, precárias as conclusões tiradas desse material acerca da região do Alto Alentejo pelos motivos acima referidos e, também, porque, para a expansão da amostra (relativamente à região indagada), se teve de recorrer a dados de épocas algo distanciadas; entretanto, alguma discussão será travada acerca do valor e eficiência das novas técnicas e de muitos dos aspectos relativos aos inquéritos por amostragem.

Finalmente, serão dadas indicações sobre o delineamento do inquérito por amostragem para o «*Plano de Fomento Agrário*» (Subsecretariado da Agricultura) — de tão decisiva importância nacional.

2. BREVE ESTUDO ESTATÍSTICO DOS RESULTADOS DO «INQUÉRITO ECONÓMICO E CULTURAL» REALIZADO EM 1930-32 À CULTURA DO TRIGO NA REGIÃO DO ALTO ALENTEJO

Para mais nítida compreensão da matéria relacionada com a moderna amostragem estatística dos inquéritos decidimos interpretar os resultados de qualquer indagação que já tivesse sido efectuada em Portugal, dispondo-nos, assim, deliberadamente a ajustar as ideias apreendidas no estudo teórico da matéria e a enfrentar as dificuldades práticas.

A fim de desvendar completamente os meandros da amostragem conviria dispor de material extenso, respeitante a alguns dos sectores da investigação. Todavia, a breve tempo demo-nos conta

dos parcos recursos nacionais quanto a inquéritos (*Cf.* DIAS Jr., 1945), pois, mesmo no campo agrário, os dados existentes eram, por sua vez, de baixo valor estatístico, por serem incompletos, de precisão e valor grosseiros, ou de restrita confiança.

No sector agrícola os únicos dados extensos e de interesse de que temos conhecimento foram publicados em 1934 pela Estação Agrária Central, e referem-se apenas à região do Alto Alentejo. Essa fraca série de elementos estatísticos (pròpriamente relativa a 1930/32) é ainda pequena; entretanto, servimo-nos dela, uma vez que dizia respeito a problemas de tão suma importância, como os económico-agrícolas.

Será sobre os resultados de tal inquérito que tentaremos aplicar as modernas técnicas estatísticas de cálculo de estimativas — na suposição de que a «amostra» constituída pelas explorações inquiridas é representativa daquela região (o nosso «universo estatístico»). Para obtermos estimativas de totais, ou médias, referentes a todo o Alto Alentejo — por expansão daquela amostra — tivémos ainda de recorrer, à falta de outros, a dados colhidos em data bastante posterior pela Junta de Colonização Interna, o que acarretou, por sua vez, mais complicações e dificuldades.

Naquela data tratara-se de ... «proceder paralelo com os trabalhos da Campanha do Trigo, a um inquérito às condições deste cereal, mormente nos aspectos económicos que o mesmo reveste», tendo-se resolvido usar na colheita de dados o «método de questionário».

Os impressos foram enviados à Lavoura e aos diversos organismos agrícolas oficiais; todavia, o inquérito ficou depois circunscrito ao Alto Alentejo, região donde se recebeu maior número de respostas, que não foi muito além de uma centena. Verificou-se mais tarde que algumas das respostas eram bastante incompletas, pelo que houve um inquérito directo complementar para suprir tais faltas.

Embora fazendo notar que «foi grande a soma de informações prestadas que teve de ser posta de parte», a Estação Agrária Central veio entretanto a publicar os resultados do inquérito respeitante àquela região (MINISTÉRIO DA AGRICULTURA, 1934), dando então igualmente conta, tanto das dificuldades encontradas para obter «com a complacência de inquiridores e inquiridos» as difíceis respostas, como dos métodos usados para tirar conclusões respei-

tantes ao estado da cultura do trigo nas explorações agrícolas do Alto Alentejo.

O interessante trabalho dedica também muitas páginas à descrição fisiográfica da região e tece ainda considerandos sobre outros problemas das explorações agrícolas, como os referentes às técnicas de cultivo, quantidade de semente usada, pessoal de trabalho, etc., muitos deles através de resumos e de tabelas, sem que os respectivos elementos base tivessem figurado na sua 1.ª Parte, a da descrição pormenorizada de cada exploração, pelo que, portanto, não podem ser incluídas no quadro que preparamos (Quadro A).

Vejamos então o que no aspecto estatístico pode apurar-se dos resultados do inquérito, particularmente dos que se referem aos atributos comuns à maioria das explorações agrícolas.

Note-se que atribuímos as «falhas» nos diversos quesitos indagados a falta de resposta (motivado tal facto por razões várias a que não é estranho o receio do empresário de que advenham novos encargos fiscais, etc.) e não à ausência da característica ou atributo nas respectivas explorações. Daí se justifica o facto de apenas se entrar em linha de conta com os casos concretos na análise estatística.

Outra nota que convém salientar aqui — por causa da «expansão» da amostra — é a seguinte: terão tratamento estatístico apenas os resultados das explorações agrícolas com 60 ou mais hectares de grandeza, visto terem sido as explorações dessas grandezas as inquiridas posteriormente por um outro organismo agrário (JUNTA DE COLONIZAÇÃO INTERNA, 1943). É esta circunstância que nos inibe de obter estimativas comparáveis com as da Estatística Oficial, o que teria o maior interesse.

Eis mais pormenorizadamente as anotações gerais a tomar em consideração; para mais fácil referência as explorações inquiridas foram por nós numeradas, seguindo-se a ordem por que figuram na 1.ª parte do trabalho da Estação Agrária Central.

Os dados de oito explorações agrícolas (as de números de ordem 6, 16, 22, 35, 39, 53, 114 e 115 da nossa catalogação) não nos merecem qualquer estudo, por serem muito incompletos. Também as explorações muito pequenas, abaixo de 60 ha, com os números 7, 18, 19, 31, 36, 55, 65, 68, 69, 70, 71, 72, 82, 84, 101, 107, 112, 113, 116, 118, 120, 127, 128 e 129 não são objecto da análise estatística geral, pelo motivo acima apontado.

Houve necessidade de fazer arredondamentos, o que é prática

usual, visto facilitar os fastidiosos cálculos sem contudo introduzir alterações de importância. Procedeu-se assim com os valores referentes à area cultural do trigo, estabelecendo-se as respectivas proporções para os dados das « despesas » e « produções ». Nestes dois casos estas práticas incidiram (logo no início da nossa interpretação) sobre as médias, expressas por ha, de cada exploração agrícola; os respectivos valores que figuram no Quadro A resultaram da multiplicação destas médias pelos valores arredondados da área de trigo. As restantes modificações são óbvias.

2 a) *Exposição sucinta de alguns métodos de tiragem de amostras e avaliação de estimativas*

A propósito dos métodos de tiragem de amostras e avaliação de estimativas é conveniente fazer ainda algumas referências gerais, mantendo-se a advertência de que se continua seguindo de perto o citado livro de YATES sobre amostragem.

Na notação das fórmulas matemáticas qualquer variável quantitativa será simbolizada por y, mas a que represente informação suplementar sê-lo-á por x.

Ainda quanto a simbolos: os dos valores da « população » serão representados por caractéres de tipo preto carregado (triplamente sublinhados, nas provas dactilografadas); os das estimativas desses valores, pelo tipo « Gill Sans » (duplamente sublinhados); e os dos valores relativos às unidades de amostragem, em itálico (simplesmente sublinhados, nas provas). Quanto às médias e totais, os símbolos correspondentes a médias levarão superiormente uma « barra » —: por exemplo, $\bar{y}$ significa a média dos valores da amostra, $\bar{\mathbf{y}}$ denota o valor da estimativa da média na população, $\bar{\mathbf{Y}}$ o valor da média na população; os respectivos totais da população são assinalados por maiúsculas.

Em quase todos os métodos de avaliação de estimativas segue-se a conhecida regra, que o autor citado enuncia da maneira seguinte: « means and proportions in the population are estimated by the corresponding means and proportions in the sample, and totals and numbers in the population are estimated multiplying the corresponding totals and numbers by the common raising factor ».

Conforme é hoje do consenso geral, a Estatística moderna fornece ainda métodos de avaliação dos *erros padrões* a que estão

sujeitas aquelas estimativas. No caso de se tratar duma amostra casual tirada duma população infinita, ou suficientemente grande, e seguindo a lei normal de erros, tem-se para a estimativa E. P. $(\bar{y})$ a generalizada equação:

$$\text{E. P. } (\bar{y}) = \frac{1}{\sqrt{n}} s$$

podendo obter-se s por intermédio da fórmula

$$s^2 = \frac{1}{n-1} S(y-\bar{y})^2 \text{ em que, } S(y-\bar{y})^2 = S(y^2) - \frac{1}{n}[S(y)]^2$$

esta última já adaptada ao uso de qualquer máquina de calcular. Note-se que no caso de populações finitas, haverá ainda que introduzir o factor $\sqrt{(1-f)}$; tal factor, todavia, não deverá ser usado em alguns casos — por exemplo, no teste de significância da diferença entre as médias de duas amostras casuais, para se verificar se provêm ou não do mesmo universo estatístico.

Não se quer entrar aqui em mais considerações, visto esta matéria figurar agora em qualquer tratado de Estatística Matemática. De passagem aponte-se apenas que há toda a vantagem em conhecer os erros padrões, pois como muito bem refere YATES: «if we know that an estimate is subject to the normal law of error and has a given standard error, we can assign probable limits of error to this estimate. If limits of plus or minus twice the standard error are taken then in 4.6 per cent of the cases will the actual error lie outside these limits». (YATES 1949 a, p. 191).

Os erros de amostragem, já aqui tantas vezes referidos, correspondem àqueles erros padrões; neles deverão estar incluídos apenas os erros casuais, a 2.ª sub-categoria de erros por nós citada logo na Introdução deste trabalho.

Quando se trata duma estimativa que seja função linear de outras estimativas, o seu erro padrão geralmente não é difícil de calcular; apenas se deverá ter em conta se estas últimas são ou não independentes (no último caso ter-se-ia de introduzir o conceito de covariância). Os diversos tipos de médias ponderadas são exemplos de funções lineares, de que também nos ocuparemos noutras alíneas. No cálculo dos erros padrões do produto e da razão de duas estimativas haverá que tomar em consideração o critério da independência.

No caso de tipos de amostras estratificadas, quase sempre o

que se pretende conhecer são os erros padrões «dentro dos estratos» e o «erro padrão generalizado», que é a média ponderada desses erros padrões. O método estatístico da análise da variância é então de aconselhar, dada a sua simplicidade. Mesmo o caso do tratamento estatístico das amostras de duas etapas torna-se, por seu intermédio, numa compreensiva aplicação (que até é simples, desde que o número de unidades de amostragem de 2.ª etapa por cada unidade de 1.ª etapa seja igual ou proporcional).

O técnico estatístico terá de esclarecer-se, para cada tipo de amostra não casual, sobre as importantes questões referentes a: «representatividade» dos estratos na população em estudo, suas grandezas relativas e, ainda, grau de variabilidade das respectivas unidades de amostragem; só depois estará habilitado a optar pelo método apropriado à obtenção de estimativas dos erros padrões das médias, ou dos totais, da população amostrada.

Resta acrescentar que a determinação da *eficiência relativa* dos diversos tipos de amostras do mesmo material permite ajuizar do mérito deles, fornecendo assim a indicação da amostra que será mais apropriada — por mais eficiente — e, depois, também a do tamanho da amostra que será susceptível de fornecer estimativas com precisão desejada.

Para ficar bem esclarecido requer este conceito de eficiência relativa, ou o seu muito próximo conceito de precisão relativa que o material do inquérito sujeito à interpretação estatística seja um tanto extenso; também haverá que atender ao critério do «custo» porque ficam ou poderão ficar os tipos de amostra propostos, pois é razão de peso que nunca deverá ser esquecida.

É de lamentar que em Portugal se não disponha de quaisquer dados concretos sobre tão importante aspecto prático dos inquéritos por amostragem. Quando muito haverá elementos respeitantes aos Censos da População no Instituto Nacional de Estatística. Salientou-se, entretanto, que deles viremos a dispor no inquérito por amostragem do «Plano de Fomento Agrário», o que poderá facilitar os delineamentos futuros, em especial os que venham a realizar-se nas nossas Províncias Ultramarinas.

Em boa verdade, na planificação de inquéritos por amostragem já hoje podem usar-se equações matemáticas para determinar a fracção óptima de amostragem, o número e a grandeza das unidades de amostra, etc. — uma vez que sejam estipulados, quer a precisão requerida, quer os custos (ainda que aproximados) de

cada uma das fases ou operações dos inquéritos. Desde há uma dezena de anos que se formularam algumas tentativas nesse sentido (*Cf.* por exemplo, JESSEN 1942) trazendo agora o livro de YATES muitos casos a exemplificar o uso de equações, em que a eficiencia é expressa em termos dos diversos componentes da variância e do custo, segundo hipóteses de valor prático.

Nas cinco sub-alíneas seguintes são descritos alguns dos principais métodos de amostragem, apresentando-se também as fórmulas de cálculo das estimativas mais importantes; pelos motivos apontados, a questão da eficiência relativa não poderá ser abordada completamente, tanto mais que apenas se dispõe de dados relativos a limitado número de explorações agrícolas.

Será, entretanto, desse material, restrito sob muitos aspectos, que viremos a obter algumas estimativas (alínea 2 b), na dupla suposição de que corresponde, sucessivamente, primeiro, a uma amostra casual e, depois, a uma amostra estratificada (aceitando que as fracções de amostragem de facto usadas por cada estrato nesta 2.ª amostra são as óptimas ou, pelo menos, as que conduzem a eficiências razoáveis).

Embora nos pareça que essa amostra não representa satisfatòriamente a região do Alto Alentejo — pois é possível que esteja afectada por algum dos «biases» da longa lista inserta no artigo «On Errors in Surveys» (DEMING, 1944), tanto mais que muitas das respostas foram obtidas por pessoas ou organismos vários não obedecendo a um critério único e bem derminado: «informadores» da Estatística Agrícola; técnicos oficiais, «categorizados» lavradores, Brigadas da Campanha do Trigo, Sindicatos Agrícolas, etc. — todavia, supô-la-emos representativa, pelas razões já apresentadas na justificação deste trabalho, perante a feição mais pròpriamente formal das conclusões que venham a tirar-se.

Iremos pois tratar, sucessivamente, os tipos de amostras que têm interesse imediato para o nosso objectivo, usando os resultados daquele inquérito de 1930/32 às explorações agrícolas (ver Quadro A) e, ainda, para permitir a «expansão» da amostra, os dados obtidos pela Junta de Colonização Interna por volta de 1943.

Expliquemos desde já a razão do emprego desses elementos estatísticos colhidos em datas algo afastadas. À primeira vista, tal prática parece errónea; todavia, é recurso que, à falta de outro

melhor, não deixa de ser muitíssimas vezes vantajoso — por permitir acréscimo de precisão das estimativas calculadas.

De facto, tal expediente é usual em inquéritos por amostragem, conforme pode ver-se através da bibliografia mundial da especialidade. Por vezes, faz parte do método de amostragem — como é o caso do uso de informação suplementar, conforme veremos mais adiante; é até recurso valioso para corrigir as inúmeras séries de elementos de pouca confiança que venham sendo publicadas rotineiramente por organismos da Estatística oficial, uma vez que se consigam em determinado momento dados rigorosos e precisos respeitantes à mesma matéria. (*Cf.* por exemplo: U. S. D. AGRICULTURE, 1949; e entre nós: VAZ, 1950).

O que se torna necessário é a descriminação das circunstâncias em que se obtiveram essas séries estatísticas, quais os critérios seguidos, as definições adoptadas, etc., etc, — tudo matéria de indagação àparte, especial e localizada, que põe mais pròpriamente à prova o juízo crítico e a responsabilidade profissional do inquiridor (DEMING, 1944).

Em cada sub-alínea refere-se, em primeiro lugar, o método de tiragem ou de selecção da amostra e, depois, o de avaliação de estimativas, com as fórmulas necessárias; por fim, dá-se um exemplo adequado. (Diga-se de passagem que não foi possível ou vantajoso adoptar no título e no conteúdo de cada sub-alínea um critério único, todavia incluir-se-á o que houver interesse em fixar).

Quanto a símbolos segue-se a notação de YATES (1949 a): $f\ (=\frac{n}{\mathbf{N}})$ corresponde à fracção prática de amostragem; n ao número de unidades de amostragem (ou unidades de tiragem) na amostra; **N** à estimativa do número de unidades na população; $g\ (=\frac{1}{f})$ ao «factor de expansão» ou «inflator»; s à estimativa de σ; **E. P.** (y) à estimativa do erro padrão da média y; etc.

O somatório S engloba as unidades da amostra; os somatórios S_i e Σ — englobam respectivamente as unidades do estrato i e do conjunto dos estratos.

2a) 1. *Amostra casual-método directo*

Uma vez definida e demarcada a população de unidades de amostragem é simples (mas nem sempre fácil) tirar a amostra desejada, devendo seguir-se um dos métodos já referidos no começo deste trabalho.

Se tal sub-divisão de todas as unidades de amostra for difícil, estabelecer-se-ão «grupos de unidades» no universo em estudo, procedendo-se

depois como para o caso do tipo de amostra em que as probabilidades de selecção são proporcionais às grandezas das respectivas unidades; nesse tipo de amostra basta apenas conhecer o número total de unidades de amostragem de cada «grupo».

Fórmulas:

Para as estimativas da média e do total na população, e dos respectivos erros padrões (caso duma variável quantitativa) (1):

$$\bar{y} = \bar{y} = \frac{1}{n} S(y) \qquad Y = g\, S(y) = N\, \bar{y}$$

$$\text{E. P.}(\bar{y}) = s \sqrt{\frac{1}{n}(1-f)} \qquad \text{E. P.}(Y) = g\, n\, [\text{E. P.}(\bar{y})]$$

Aplicação:

Podemos exemplificar com a variável «área de trigo» para $n = 72$ explorações agrícolas de 60 ou mais hectares, aquelas em que há, conjuntamente, elementos estatísticos para «despesas com a cultura do trigo» e para «produção de trigo». Atendendo a que: (Quadro A, parte inferior) $460 + 210 + \ldots + 810 = 8.440$ e $32.600 + 7.300 + \ldots + 178.900 = 2.116.400$, e, ainda, a que $N = 1803$ (*Cf.* JUNTA DE COLONIZAÇÃO INTERNA, 1943, ou Tabela III) e, como consequência, $f = \frac{72}{1.803} = 0{,}039933$ e $g = 25{,}042$ — por substituição nas fórmulas apropriadas já referidas anteriormente tem-se:

$$S(y-\bar{y})^2 = 2.116.400 - \frac{1}{72}(8.440)^2 = 1.127.040, \qquad s^2 = \frac{1.127.040}{71} = 15.874$$

vindo, então, para as estimativas requeridas os seguintes valores:

$$\bar{y} = 8.440/72 = 117{,}0 \qquad Y = 211.350$$

$$\text{E. P.}(\bar{y}) = \sqrt{15.874\left(1 - \frac{1}{25{,}042}\right)/72} = \pm 14{,}55 \qquad \text{E. P.}(Y) = \pm 26.250$$

que figuram na Tabela III da alínea 2 b) deste trabalho. Os valores correspondentes para a «área de trigo» ou a «grandeza total» das 97 explorações agrícolas de 60 ou mais ha, que foram inquiridas, podem ser obtidos de modo análogo. Para o primeiro caso tem-se, por exemplo:

$$Y = 222.670 \pm 22.950$$

Note-se que, quando **N** é conhecido, como no presente caso, então subtituir-se-á a N nas fórmulas anteriores.

---

(1) Para proporções (sendo u o número de unidades correspondendo, na amostra, a determinado atributo) ter-se-ia, para erros padrões de $p = u/n$ e $U = N\,p$, respectivamente:

$$\text{E. P.}(p) = [p\,q\,(1-f)/n] \quad \text{e,} \quad \text{E. P.}(U) = N\,[\text{E. P.}(p)]$$

Há ainda outra modalidade deste tipo de amostra, conhecida por «*amostra casual, estratificada depois da selecção*», que permite eliminar as diferenças entre os estratos. Com tal amostra é necessário conhecer a grandeza de cada um dos estratos na respectiva população. É conveniente que a amostra não seja pequena, a fim de que haja um número elevado de unidades de tiragem em cada estrato e fique portanto assegurada a respectiva «representatividade»; uma vez que estas condições sejam respeitadas, faz-se o cálculo das estimativas segundo algum dos métodos aplicáveis às amostras estratificadas.

2a) 2. *Amostra casual — método da regressão (e da «razão»).*

Viu-se no começo deste trabalho que o uso da informação suplementar nos inquéritos por amostragem é um dos valiosos meios de aumentar a precisão das estimativas.

Dos dois métodos de avaliação de estimativas: o da regressão e o da razão — o primeiro é já largamente conhecido entre nós; o segundo é bastante semelhante ao primeiro, e tem a vantagem de exigir cálculos menos fastidiosos. Segundo YATES (Ibid., p. 156): «the regression method therefore differs from the ratio method in that in the former the straight line which best fits the sample values is taken, whereas in the later the line through the origin is taken».

Para se aplicarem esses métodos basta conhecer o total ou a média da respectiva variável x na população (ou nos estratos). No método da regressão linear — que é suficiente nos casos gerais — calcula-se primeiramente o coeficiente de regressão na amostra estatística, o qual será depois usado para «to adjust the sample results for any discrepancy between the mean size of units in the sample and in the population» (YATES *op. cit.*, p. 155). O método da razão será tratado mais desenvolvidamente noutra sub-alínea, embora se trate então do caso duma amostra estratificada.

Qualquer dos referidos métodos tem particular importância nos inquéritos económico-agrícolas por amostragem, visto que é relativamente fácil conhecer o total (a área de todas as explorações agrícolas ou, pelo menos, a área agrícola dum país, ou duma região, etc.) da variável suplementar, através de inquéritos gerais e expeditos.

No inquérito agrário do Alto Alentejo levado a efeito em 1943 pela Junta de Colonização Interna, conclui-se que a área total das explorações agrícolas (de 60 ou mais hectares) media 1.061.051 ha na Carta Agrícola de 1/50.000. Ora nas $n = 97$ explorações inquiridas em 1930/32 — Quadro A — o somatório das respectivas áreas totais atinge o valor de 84.780 ha, enquanto o da área de trigo é de 12.000 ha.

Por aplicação de fórmulas apropriadas à amostra de 97 explorações agrícolas as estimativas que se obtiveram para a área de trigo foram: $y = 123{,}7 \pm 11{,}09$ (ou $\mathbf{Y} = 191.060 \pm 19.995$ ha) com o método da regressão; e $y = 83{,}3 \pm 9{,}28$ (ou $\mathbf{Y} = 150.180 \pm 16.730$ ha,) com o método da razão.

Conforme pode observar-se, a estimativa do total apresenta agora um valor bastante baixo, todavia está afectada de mais alta precisão.

Os métodos da regressão e da razão são muito vantajosos nas amostras contrabalançadas (« balanced samples »), e na denominada amostragem por « fases múltiplas », de tanto interesse para nós, principalmente em relação às nossas Províncias Ultramarinas. O seu papel na correcção das séries estatísticas oficiais reveste-se também da máxima importância, conforme se referiu.

2a) 3. *Amostra estratificada (fracção constante de amostragem) — método directo*

Quando o universo a estudar por amostragem é heterogéneo, como acontece nos casos agrícolas, e é possível dividi-lo em grupos ou estratos relativamente homogéneos — usa-se então, muitas vezes, a denominada amostra estratificada, porque dá acréscimo de eficiência.

Adoptando uma mesma fracção constante de amostragem, de cada estrato tirar-se-á uma amostra, para o que se seguirá o método de selecção indicado na sub-alínea 2a) 1, no caso de ser casual; o conjunto de tais amostras dos estratos é que constitui a amostra desejada.

Fórmulas:

Com este tipo de amostra estratificada o cálculo das estimativas (pelo método directo) faz-se separadamente, estrato a estrato, devendo então aplicar-se as fórmulas citadas para o caso da amostra casual.

Uma vez obtidas as estimativas das variâncias das médias ou totais de cada estrato — que serão tanto mais estáveis quanto maiores forem as grandezas dos estratos — a estimativa da variância da média ou total da população respectiva será fàcilmente expressa como função linear daquelas. Como apenas se dispõe de poucos dados, não se fará a exemplificação deste caso de amostragem.

Note-se que com tal tipo de amostra ainda pode ser usado o método da razão (ou o da regressão) para avaliar as estimativas de médias e de totais, quer empregando uma mesma razão para todos os estratos, quer diferentes razões, conforme estes diversos grupos.

2a) 4. *Amostra estratificada (razão variável de amostragem) — método directo*

Já atrás se referiu que este tipo de amostra é da mais alta importância no sector agrário, onde o material estatístico varia consideràvelmente, visto permitir obter estimativas de precisão muito elevada.

De facto, quando se classificam as explorações agrícolas segundo a grandeza logo se reconhece que as explorações maiores são precisamente as que manifestam maior variabilidade; por isso, as fracções de amostragem, f_i, que pelo princípio da « distribuição óptima das unidades da amostra » (NEYMAN, 1943) devem ser proporcionais a σi, são simplesmente tomadas em proporção às grandezas médias dos estratos, as classes de grandeza — o que simplifica a questão do delineamento dos inquéritos por amostragem.

Já anteriormente nos referimos a esta matéria (OLIVEIRA 1948, cap. 4),

tendo-se visto que no «National Farm Survey of England and Wales» se usaram fracções de amostragem desde 1/20, 1/10...a 1/1 para os estratos de explorações respectivamente com a grandeza 2-10, 10-40,... e acima de 280 ha; resultava desse facto incluir-se na amostra quase todas as explorações agrícolas de maior «peso» económico, uma das maiores vantagens práticas do grande inquérito.

Entre nós conviria realizar um inquérito que permitisse dar a conhecer os diversos factores inerentes às questões agrárias nacionais (para o que se recomenda um estudo prévio e rápido, género de pesquisa de orientação — «pilot survey» — que, por sua vez, permitiria ajustar verbas e treinar pessoal técnico).

Vejamos entretanto, o que se pode extrair do material estatístico do inquérito da antiga Estação Agrária Central, exemplificando com a variável «área de trigo» da amostra constituída por 72 explorações agrícolas, mas agora supostamente estratificada em 3 classes de grandeza (1), conforme se mostra na Tabela I.

Não se dispõe de dados abundantes e, por isso, as fracções de amostragem serão aquelas que de facto o material disponível comportar e não as correspondentes à distribuição óptima da amostra.

Fórmulas:

As fórmulas a seguir indicadas para avaliar as estimativas dos totais no universo estatístico não apresentam quaisquer dificuldades na sua aplicação, apenas os cálculos devem ser feitos estrato a estrato:

$$\mathbf{N} = \Sigma\,(g_i\, n_i) \qquad\qquad \mathbf{Y} = \Sigma\,[g_i\, S_i\ (y)]$$

Para o cálculo do erro padrão das estimativas determinam-se as respectivas variâncias também estrato a estrato, adicionando-as depois, uma vez ponderadas. YATES no livro já referido exemplifica a dedução da variância da estimativa da média $\mathbf{y} = \Sigma\,[g_i\, S_i\,(y)]/\mathbf{N}$ respeitante ao universo estatístico, chegando ràpidamente à fórmula:

$$\mathbf{V}\,(\bar{\mathbf{y}}) = \Sigma\,[g_i^2\ \ n_i \sigma_i^2\ \ (1\text{-}f_i)]/\mathbf{N}^2$$

uma vez que $\mathbf{V}\,[S_i\,(y)] = n_i \sigma_i^2\,(1\text{-}f_i)$, e σ_i^2 corresponde à variância dentro de cada estrato. (No cálculo de $\mathbf{V}\,(\bar{\mathbf{y}})$ os σ_i^2 serão substituídos pelas suas estimativas s_i^2).

Note-se que, ao passo que podia fazer-se uso da técnica da análise da variância no caso duma amostra estratificada com fracção constante de amostragem, do tipo simples «entre» e «dentro» dos estratos, agora tal não pode suceder — e, por conseguinte, aquelas estimativas deverão ser obtidas estrato a estrato, poderando-as com pesos proporcionais a $g_i^2\, n_i\ \ (1\text{-}f_i)$.

(1) Grandezas médias das 122 explorações inquiridas, nos 4 estratos: 17,1; 132,5; 582,0 e 2.211,5 ha, em número respectivamente de: 25, 24, 49 e 24 (o 1.º estrato inclui as explorações com menos de 60 ha de grandeza).

Aplicação:

Na Tabela I incluem-se as fases principais do cálculo das estimativas da «área de trigo», respeitantes às explorações agrícolas com 60 e mais hectares de grandeza, e dos seus respectivos erros padrões — valores que virão a figurar na Tabela III da alínea seguinte.

TABELA I

Avaliação de estimativas de totais e médias e seus erros padrões, da «área de trigo» de explorações agrícolas do Alto Alentejo. Amostra estratificada (fracção variável de amostragem) — método directo

Classes de grandeza	N.º de explorações		Factor de expansão	Área de trigo na amostra		$s_i^2 = \frac{S_i(y-\bar{y})^2}{(n_i-1)}$	E. P. $(\bar{y}_i) = \sqrt{\frac{s_i^2}{n_i}(1-f_i)}$	$V(S_i(y)) = n_i s_i^2 (1-f_i)$
	N_i (*)	n_i	g_i	total	média			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
60- 250	716	19	37,684	670	35,5	905	± 6,81	16.740
250-1000	824	33	24,970	4.000	121,0	15.850	± 21,47	502.130
1000-	263	20	13,150	3.770	188,5	19.675	± 30,15	363.590
—	1803	72	25,042	8.440	—	—	± 11,09	—

Para se obter a estimativa da área total de trigo na região de Alto Alentejo (das explorações com 60 ou mais hectares) torna-se necessário calcular, primeiramente, as estimativas para cada estrato, bastando multiplicar os valores da coluna 4 pelos correspondentes da coluna 5 da Tabela I; então adicionando os produtos obtidos tem-se o valor de Y, ou seja no caso concreto Y = 174.710 ha, e, depois, a média y = 97,0 ha (*Cf.* Tabela III).

Quanto ao erro padrão da estimativa do total na população em estudo, em primeiro lugar, a variância do total da «área de trigo» de cada classe ou estrato (cada valor da coluna 9 da Tabela I) deve multiplicar-se pelo quadrado do respectivo «factor de expansão» g_i; seguidamente, adicionam-se as variâncias assim obtidas, devendo extrair-se depois a raiz quadrada à soma obtida, que no nosso caso é 399.721.000.

O erro padrão requerido tem, pois, o valor $\sqrt{399.721.000} = \pm 19.990$ ha; à média y = 97,00 corresponde o erro padrão 19.990/1.803 $\Rightarrow \pm 11,09$. O último valor coincide com o que pode obter-se directamente por aplicação da fórmula $\sqrt{V(\bar{y})}$.

Deve acrescentar-se que não é estritamente necessário usar fracções

(*) Os elementos desta coluna foram cedidos pela Junta de Colonização Interna, pelo que lhe deixamos expressos os nossos agradecimentos.

de amostragem com décimais, visto que não se aumenta consideràvelmente a precisão das estimativas mas sòmente torna mais fastidiosos os cálculos.

2a) 5. *Amostra estratificada (fracção variável de amostragem) — método da razão*

No aspecto económico-agrícola tem particular interesse a determinação do *« preço de custo de 1 kg de trigo »*, que, estatisticamente, mais não é do que um dos casos em que se faz uso da denominada informação suplementar.

A fim de facilitar a compreensão da matéria relacionada com aquela determinação vamos indicar os principais aspectos do método da razão para o cálculo de estimativas das médias e totais no universo em estudo, usando primeiramente a *amostra casual* das 72 explorações agrícolas inquiridas em 1930/32, já anteriormente referida.

A variável suplementar x corresponderá à característica « produção de trigo » das explorações agrícolas e y representará a variável « despesas com a cultura do trigo ». Note-se que os atributos são considerados por exploração e não por hectare, entrando-se assim já em consideração com o factor « grandeza » das explorações agrícolas.

Segundo as próprias palavras de YATES, p. 155 do seu livro já referido:

... « In the ratio method, the ratio of $\mathbf{Y}/\mathbf{X}$ in the population is estimated from the sample, the estimated ratio being multiplied by the total $\mathbf{X}$ of x for the population to give the estimated total $\mathbf{Y}$ of y for the population. The method of estimation must be such that bias is avoided ».

... « The appropriate estimate of the ratio for a random sample is $S(y)/S(x)$ or $\bar{y}/\bar{x}$ ».

Pela transcrição dada fica-se já de posse das fórmulas apropriadas à avaliação das estimativas r e Y do universo, faltando-nos apenas indicar as relativas aos erros padrões respectivos.

A fórmula de $\mathbf{V}(\mathbf{r})$ corresponde à da variância da razão de duas estimativas devendo apenas incluir-se-lhe o termo da covariância (dado que os valores de x e y pertencem a cada unidade de tiragem), e o termo $1-f$ (por causa do carácter finito da população que se considera).

Tem-se pois:

$$\mathbf{V}(\bar{\mathbf{r}}) = \frac{1-f}{n}\,\bar{\mathbf{r}}^2\left(\frac{\mathbf{V}(y)}{\bar{y}^2} - \frac{2\,\mathbf{cov}(x\,y)}{\bar{x}\,\bar{y}} + \frac{\mathbf{V}(x)}{\bar{x}^2}\right)$$

Note-se que esta fórmula é aproximada, mas pode tomar-se como certa para a maioria dos casos correntes: pode ainda dar-se-lhe a seguinte expressão prática:

$$\mathbf{V}(\bar{\mathbf{r}}) = \frac{1-f}{n\,\bar{x}^2}\,s_q^2 = \frac{1-f}{[S(x)]^2}\,n\,s_q^2$$

em que $s_q^2 = Q/(n-1)$, com $Q = S(y-y)^2 - 2\,\mathbf{r}\,S(x-x)(y-y) + \bar{\mathbf{r}}^2\,S(x-x)^2$ representando Q a soma de quadrados dos desvios dos valores de y em relação aos valores dados pela « linha da razão ».

Quanto à variância de Y usa-se a expressão matemática:

$$V(Y) = \frac{X^2}{[S(x)]^2} (1-f)\, n\, s_q^2$$

No caso concreto da amostra casual de 72 explorações agrícolas representando a região do Alto Alentejo nas condições já referidas, e para a qual é: $n = 72$, $1-f = 0{,}9601$ e, ainda (ver o Quadro A):

$\bar{r} = S(y)/S(x) = \$8895$ $\quad S(y-\bar{y})^2 = 315.382.960.000$

$S(x) = 5.463.500$ $\quad S(x-\bar{x})(y-\bar{y}) = 284.158.510.000$

$S(y) = 4.860.000\$00$ $\quad S(x-\bar{x})^2 = 396.526.430.000$

tem-se, depois das devidas substituições na fórmula anteriormente referida:

$$V(\bar{r}) = \pm \$004032, \text{ donde } E.P.(\bar{r}) = \pm \$06351,$$

precisamente o valor que figura na Tabela III. Diga-se de passagem que nalguns casos podem usar-se (com outras fórmulas apropriadas) os valores de r de cada unidade de amostra; tal método, todavia, requer atenção especial, como seja a de verificar, por exemplo, se os valores de r estão sujeitos a qualquer marcha ou tendência («trend»).

Vejamos agora pròpriamente o caso do cálculo do «preço de custo de 1 kg de trigo» numa *amostra estratificada*, em que se use o método da razão como método para avaliar as estimativas do universo estatístico.

Duas hipóteses podem ser consideradas: a) — uso duma mesma razão para todos os estratos; b) — uso de razões diferentes, conforme os estratos. O 2.º caso é mais trabalhoso e deve ser empregado, de preferência, quando os estratos da população estejam bem representados pelos correspondentes estratos da amostra.

Isto não acontece por certo com o material estatístico que vimos analisando; contudo, e a título puramente de aplicação das fórmulas anteriores a cada estrato, apresentamos a seguir os resultados obtidos estrato a estrato, por cada base de estratificação (Tabela III) — na amostra de 72 explorações:

1) Grandeza da exploração

Classes	$\bar{r}_i$	$\pm$	E.P. $(\bar{r}_i)$
60- 250	\$92,47	$\pm$	\$05,552
250-1000	\$78,64	$\pm$	\$09,841
1000- . .	\$99,75	$\pm$	\$07,298
(geral)	(?)		(?)

2) Forma de exploração

Categorias	$\bar{r}_i$	$\pm$	E.P. $(\bar{r}_i)$
Conta propria	\$84,87	$\pm$	\$09,402
Arrendamento	\$91,24	$\pm$	\$06,550
Mista	1\$04,29	$\pm$	\$09,077
(geral)	(?)		(?)

Os casos assinalados com (?), que correspondem aos das determinações dos valores de r e seus erros padrões pelas fórmulas $r = Y/X$ e $V(\bar{r}) = V(Y)/X^2$, em que:

$$Y = \Sigma \left(\frac{S_i(y)}{S_i(x)} \mathbf{X}_i \right)$$

$$V(Y) = \Sigma \left(\frac{\mathbf{X}^2}{[S_i(x)]^2} (1-f_i) n_i s^2_{q_i} \right),$$

significam impossibilidade, visto que não se conhecem nem os valores de $\mathbf{X}_i$ (e portanto de $\mathbf{X}$), nem os da produção total de trigo para cada estrato — elementos não inquiridos ou pelo menos não publicados quer, no seu tempo, pela Estação Agrária Central ou outro organismo superior, quer, mais modernamente, pela Junta de Colonização Interna.

Seria do mais alto valor agrário determinar os preços de custo do trigo, segundo as classes de grandeza e as formas de exploração, mesmo com os dados de um único ano agrícola; mas acabamos de ver que o material estatístico é bastante incompleto, pelo que não convem, pois, entrar em estudo minucioso, como seria, por exemplo, o relacionado com as provas de significância das diferenças entre as médias, r_i, dos diversos estratos (*).

* * *

Resta-nos focar o aspecto *«eficiência relativa»* dos diferentes métodos de amostragem, questão que se levanta sempre que se tenha de delinear qualquer inquérito por amostragem, em particular se se quizerem considerar os diversos gastos monetários e de tempo nas várias fases de execução dos trabalhos.

É já de todos conhecida a fórmula do desvio padrão da média — mais simplesmente designado por «erro padrão» para o caso duma amostra casual dum grande universo — e que é assim expressa (com os desvios padrões em percentagem das médias):

$$\text{erro padrão de } (\bar{y}), \text{ em } \% = \frac{\sigma\,\%}{\sqrt{n}}$$

substituindo-se correntemente σ^2 por s^2, sua melhor estimativa. Pois bem, tal fórmula expressa destrouto modo:

$$n = \frac{(\text{desvio padrão, em } \%)^2}{(\text{erro padrão requerido, em } \%)^2}$$

permite obter a grandeza da amostra que conduz a determinada

(*) Outros métodos de amostragem poderiam ser seguidamente exemplificados: todavia, para não alongarmos demasiado o presente trabalho não o faremos aqui. Para elucidação da matéria deve consultar-se o livro de YATES (1949 a) em especial para o estudo dos importantes métodos de amostragem de: «fases múltiplas» e «etapas múltiplas». O caso da «amostragem sistemática» é também aí tratado com algum desenvolvimento.

precisão, uma vez que seja já conhecido o desvio padrão duma outra amostra casual.

Vejamos um exemplo concreto com o caso da variável «área de trigo» das 97 explorações agrícolas de 60 ou mais hectares de grandeza inquiridas pela antiga Estação Agrária Central, e constituindo uma amostra, suposta casual, das correspondentes explorações agrárias do Alto Alentejo.

Na sub-alínea 2 a) 2, tinha-se $s^2 = \frac{1.629.070}{(97-1)} = (\pm 130)^2$; como a estimativa $y = 123{,}7$ ha, vem, pois, desvio padrão em $\% = \frac{130 \times 100}{123{,}7} = 105\,\%$.

Se agora se quizer saber qual a grandeza verdadeira, ou pelo menos aproximada, da amostra de explorações agrícolas que conduza a estimativas com um erro padrão de 7 %, o emprego da fórmula citada diz-nos que:

$$n = \frac{(105)^2}{(7)^2} = 225$$

enquanto que seriam necessárias 441 explorações para fazer baixar o erro padrão para 5 %.

Se se atender ao facto de que a população é finita, tem de se fazer uma correcção. O número acabado de determinar corresponderá a n_o; depois calcula-se $f_o = \frac{n_o}{N}$. Então a fracção de amostragem, para se conseguir a precisão desejada, será dada por: $f = f_o/(1 + f_o)$. Nos exemplos apontados os valores de n recalculados por uso desta fórmula são 200 e 330, respectivamente.

Com as amostras casuais é, pois, muito fácil o cálculo de n. Todavia, surgem algumas dificuldades quando o material básico corresponde a amostras doutros tipos, conforme se pode notar nos exemplos que vão ser dados, em que se usa o conceito da precisão relativa, diferente do da eficiência relativa, que são definidos por YATES (*op. cit.*, p. 247) da maneira seguinte: «The relative precision of two different methods of sampling based on the same type of sempling unit may be defined as the reciprocal of the ratio of the sampling variances of the estimates given by the two methods when the same number of units are taken», e «The relative efficiency... may be defined by the reciprocal of the ratio of the numbers of units required to attain a given accuracy with the two methods».

Note-se que é possível alargar o âmbito destes conceitos mesmo a casos de amostragem com diferentes tipos de unidades

de amostra, ou em que se faça intervir o custo, etc. Nos exemplos que apresentamos, tirando ainda algum partido do material estatístico do inquérito de 1930/32 à cultura do trigo, apenas nos reportaremos, todavia, aos casos simples e imediatos.

1) Cálculo da precisão relativa entre a «amostra estratificada com fracção variável de amostragem — método directo» e a «amostra casual — método directo» a partir dos dados da área de trigo duma amostra estratificada, com fracção variável de amostragem, de explorações do Alto Alentejo.

Para o caso em referência usar-se-ão as seguintes fórmulas (aliás relacionadas ainda com a análise da variância duma «amostra estratificada — fracção constante de amostragem»):

$$s_1^2 = \Sigma h_i s_i^2$$

$$s^2 = s_1^2 + \Sigma h_i \bar{y}_i^2 - \bar{\mathbf{y}}^2 - \Sigma h_i (1-h_i) s_i^2/n_i$$

em que as proporções h_i têm por fórmula: $h_i = \mathbf{N}_i / \mathbf{N}$. A estimativa da média na população é obtida a partir da amostra, usando a expressão matemática: $\bar{\mathbf{y}} = h_i \bar{\mathbf{y}}_i$.

Os cálculos necessários figuram na Tabela II. Os valores das colunas 3, 4 e 6 são os correspondentes das colunas 3, 6 e 7 da Tabela I que inclui a «amostra estratificada (fracção variável de amostragem) — método directo». O valor 12.378,8 da coluna 5 resulta da aplicação da fórmula que figura em 2.º lugar na expressão da direita para o cálculo de s^2.

Para a determinação da precisão relativa tem-se agora (sendo $96{,}91^2 = 9391{,}5$):

$$s_i^2 = 10.475{,}3$$

$$s^2 = 10.475{,}3 + 12.378{,}8 - 9.391{,}5 - 253{,}3 = 13.209{,}3$$

e vem finalmente: $13.209{,}3 / 10.475{,}3 = 1{,}26$ o que significa um ganho de precisão de 26 %. Note-se que no exemplo citado as correcções da coluna 7 são relativamente pequenas.

Se se quiser saber a *grandeza* da amostra que conduz, por exemplo, a um erro padrão de $\pm$ 10.000 ha para a estimativa do

TABELA II

Cálculo da «variância média dentro dos estratos» e da «variância geral», a partir da «amostra estratificada (fracção variável de amostragem) — método directo» de 72 explorações agrícolas do Alto Alentejo.

Classes de grandeza	h_i	n_i	$\bar{y}_i$	$\bar{y}^2_i$	s^2_i	$h_i(1-h_i)\, s^2_i / n_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
60-250	0,397	19	35,5	1.260	905	11,4
250-1000	0,457	33	121,0	14.641	15.850	119,2
1000- .	0,146	20	188,5	35.532	19.675	122,7
—	1,000	72	96,91	12.378,8	10.475,3	253,3

total da «área de trigo» usando as mesmas fracções de amostragem da Tabela I (em números redondos, respectivamente 1/38, 1/25 e 1/13 para as três classes de grandeza), é necessário recorrer à seguinte fórmula, indicada por YATES (*Ibid.*, p. 255):

$$\frac{1}{c} \Sigma (s^2_i\, h_i / \lambda_i) = \mathbf{N}\, \mathbf{V}(\bar{y}) + \Sigma\, s^2_i\, h_i$$

na qual já conhecemos o significado de muitos dos símbolos. Nesta fórmula $f_i = c\, \lambda_i$; c deve ser calculado de acordo com a precisão requerida.

Com os dados da Tabela II tem-se (correspondendo λ_i às aludidas fracções de amostragem):

$$\Sigma (s^2_i\, h_i / \lambda_i) = 232.082 \text{ e, } \Sigma\, s^2_i\, h_i = 10.475{,}3$$

e, ainda:

$$\mathbf{N}\, \mathbf{V}(\bar{y}) = \mathbf{V}(y)/\mathbf{N} = (10.000)^2/1.803 = 55.463$$

Por conseguinte, vem $c = (232.082)/(55.463 + 10.475{,}3) = 3{,}52$.

A amostra requerida deverá, pois, comportar: $72 \times 3{,}52 = 253$ explorações agrícolas pelo menos (de 60 ou mais ha de grandeza total): a sua primeira classe de grandeza incluirá $716 \times 3{,}52/38 = 66$ explorações, a 2.ª classe: $824 \times 3{,}52/25 = 116$ explorações, etc.

Mais uma vez salientamos o facto de ser de pouca confiança, quanto a «representatividade», o material que vimos analizando

estatisticamente, pelo que as sugestões obtidas tem de ser consideradas com reservas. De resto, YATES diz peremptòriamente: «No very general conclusions should, however, be drawn from a single comparison based on a small amount of data even when a single set of data is used» (*Ibid.* p. 247).

2) Estudo dos «ganhos de precisão» no cálculo da estimativa da área de trigo duma amostra casual de explorações agrícolas do Alto Alentejo, por uso de informação suplementar (método da regressão).

O ganho de precisão aparente é calculado, aproximadamente, pela simples proporção dos quadrados dos erros padrões das amostras a comparar.

Nos exemplos concretos mencionados é possível verificar (*Cf.* 2 a 1) que a estimativa da área total de trigo, a partir da «amostra casual — método directo» de 97 e xplorações agrárias, estava afectada dum erro padrão com o valor $\pm$ 22.950, enquanto que a da «amostra casual — método da regressão» (sub-alínea 2 a 2) tinha por erro padrão $\pm$ 19.995.

Por conseguinte, o ganho de precisão devido ao uso da variável suplementar no referido método da regressão é: $(22.950)^2/(19.995)^2 = 1{,}32$. Se se exprimir a variância da «amostra casual — método directo» por 100, a variância relativa, por exploração, correspondente à «amostra casual — método da regressão» será expressa por 76, o que significa serem 100 e 76, respectivamente, os números de explorações agrícolas necessários para alcançar a mesma precisão.

Com o método da razão (ver a mesma sub-alínea 2 a 2) ter-se-ia, análogamente: $(22.950)^2/(16.730)^2 = 1{,}88$.

Ora o método da regressão é, pelo menos, tão eficiente como o da razão, e daí o dever procurar-se um mais perfeito processo de cálculo da precisão relativa. Usando a fórmula: $1/(1-r^2)$, para o 1.º método, em que r corresponde ao conhecido coeficiente de correlação (no caso presente com o valor 0,5143) e estoutra:

$$\frac{1}{1-r^2+r^2\left(1-\frac{\bar{r}}{b}\right)^2}$$

para o 2.º método, na qual agora é $\bar{r} = S(y)/S(x)$, e b o coeficiente de regressão, fazendo as substituições apropriadas obtêm-se

os seguintes valores para a precisão relativa, respectivamente: 1,36 e 0,86. Conforme se nota, uma vez que haja estimativas correctas da precisão relativa, o método da regressão revela-se como o mais eficiente.

O cálculo da *grandeza* da amostra em face duma dada precisão também não oferece quaisquer dificuldades, uma vez que se sigam os métodos de determinação de variâncias indicados em 2 a 2.

2 b) *Análise dos resultados do inquérito a explorações agrícolas do Alto Alentejo. O caso do «preço de custo» do trigo*

Nesta altura do trabalho podemos mais à vontade interpretar os resultados do «inquérito económico e cultural» de 1930/32, alguns dos quais temos já utilizado na exemplificação de alguns métodos de amostragem e cálculo de estimativas.

Devido à exiguidade do material estatístico para muitas das características das explorações agrícolas inquiridas, apenas nos referiremos aos resultados de 5 dessas variáveis, aliás as mais importantes. Pelo motivo já anteriormente apontado no início do Capítulo II, sòmente merecem aqui atenção particular as empresas de 60 ou mais ha de grandeza (Quadro A).

No estudo de cada uma das 5 variáveis em questão supor-se-à que houve tiragem prévia de duas amostras independentes, uma casual, I, e outra estratificada, II (*Cf.* Tabela III); será sobre essas amostras que incidirá a análise estatística, adoptando-se como método de avaliação das estimativas dos universos respectivos o *método directo.*

Deixar-se-á para a alínea 3 a) a questão da «representatividade» dessas amostras de explorações agrícolas, que é problema da mais alta importância nos inquéritos por amostragem, e precisa, por isso, de atenção especial; o problema da base de estratificação, ainda relacionado com este, será também um tanto discutido nessa altura.

Vejamos o caso das três características para as quais o inquérito forneceu conjuntamente dados: «área de trigo» (em ha), «despesas com a cultura» (em escudos) e «produção de trigo» (em kg).

Por aplicação dos métodos descritos nas sub-alíneas 2 a 1 e 2 a 4 aos elementos que figuram ordenadamente no Quadro A — em cuja parte inferior se inscreveram alguns resultados do cálculo de totais e «somas de quadrados» — obtiveram-se as estimativas

QUADRO A

Classificação dos dados do « Inquérito económico e cultural » realizado em 1930/32 à cultura do trigo no Alto Alentejo (Cf. MINISTERIO DA AGRICULTURA, 1934) em explorações com 60 ou mais ha — segundo a « Grandeza da exploração » e a « Forma de exploração ».
(Os valores que figuram em cada « célula » dizem respeito às características a), b), c), d) e e) das explorações).

60 — 250 ha		250 — 1000 ha			1000 — ... ha		
Conta própria	Arrendamento (total)	Conta própria	Arrendamento (total)	Mista	Conta própria	Arrendamento (total)	Mista
50 26 38.000$	30 4 24.600$	40 28 26.800$	70 27 49.700$	60 15 —	80 17 48.800$	50 55 31.500$	140 26 30.800$	100 32 104.000$
50.500 370	21.600 —	— —	79.100 530	— —	49.600 —	86.500 1.110	96.600 660	81.000 1.650
100 — —	40 7 11.200$	100 44 120.000$	80 — —	80 65 69.600$	180 40 —	80 50 45.600$	650 — —	200 39 84.000$
71.000 100	13.600 —	100.000 420	77.600 690	76.000 300	138.600 —	39.200 —	— 540	70.000 1.430
40 — —	20 3 12.000$	180 50 75.600$	120 50 —	150 33 66.000$	290 — —	120 55 78.000$	150 21 111.000$	310 40 155.000$
39.600 250	16.600 70	104.400 850	72.000 —	84.000 1.310	272.600 1.540	121.200 1.000	85.500 990	111.600 —
40 8 24.800$	10 5 5.600$	660 50 66.000$	90 37 —	100 20 47.000$	250 43 147.500$	270 105 205.200$	600 150 240.000$	180 22 79.200$
23.200 —	7.300 210	382.800 520	63.000 —	37.000 —	142.500 —	162.000 —	300.000 4.840	127.800 1.640
30 32 14.400$	30 4 18.300$	80 15 —	120 21 —	240 — —	150 27 34.500$	80 27 72.000$	280 52 182.000$	360 21 —
11.700 —	13.800 120	— —	— —	240.000 610	52.500 400	68.800 —	196.000 —	— —
10 5 7.600$	50 4 24.500$	80 50 52.800$	70 32 —	110 — —	20 — 17.200$	420 60 373.800$	80 21 62.400$	20 45 18.600$
9.400 30	38.500 —	79.200 480	— —	49.500 860	18.600 20	277.200 —	56.000 1.700	17.200 —
10 5 3.800$	50 35 —	130 — —	80 22 56.000$	70 14 37.800$	50 6 40.500$	80 — 60.000$		
7.500 80	35.000 —	123.500 1.190	28.800 870	44.800 1.150	28.500 650	60.000 150		
10 2 8.200$	30 6 18.900$	60 40 34.200$	120 — 90.000$	50 10 19.500$		280 — 193.200$		
7.400 40	16.800 —	38.400 —	78.000 1.120	36.500 150		198.800 —		
30 1 22.200$		90 15 71.100$	100 25 82.000$	100 32 110.000$		340 30 —		
30.900 —		70.200 1.030	171.000 1.420	86.000 790		238.000 —		
10 30 —		50 32 —	60 47 49.200$	160 55 84.800$		150 36 103.500$		
7.000 —		— —	64.200 1.520	68.800 400		75.000 760		
30 12 —		10 15 4.400$	90 — —	50 17 41.000$		200 30 —		
— —		7.700 —	134.100 1.550	36.000 160		— —		
140 60 86.800$		80 6 21.600$	200 6 40.000$	40 5 33.200$		180 5 138.600$		
98.000 —		33.600 —	56.000 —	38.800 1.800		144.000 —		
70 7 29.400$		100 36 62.000$	450 22 270.000$					
30.100 350		73.000 1.310	310.500 —					
10 5 11.800$		30 31 —	80 — 55.200$					
11.700 350		— —	67.200 —					
30 2 12.300$		100 — 57.000$						
12.900 70		75.000 150						
30 — 42.000$		30 47 21.000$						
28.800 —		20.100 160						

Grandeza da exploração	Forma de exploração	a) Área anual de trigo n	$S(y)$	$S(y^2)$	b) Pessoal certo n	$S(y)$	$S(y^2)$	c) Despesas com o trigo n	$S(y)$	$S(y^2)$	d) Produção de trigo n	$S(y)$	$S(y^2)$	e) Gado (geral) n	$S(y)$	$S(y^2)$
60-	C / própria	12	460	32.600	13	195	6.541	12	301.300$	13.351.810.000$	12	322.100	16.022.310.000	9	1.640	468.200
-250	Arrend. (total)	7	210	7.300	8	68	1.392	7	115.100$	2.198.310.000$	7	126.200	2.935.300.000	3	400	63.400
250-	C / própria	19	2.650	811.700	24	748	27.922	19	1.277.800$	141.901.740.000$	19	1.839.200	343.251.680.000	16	13.810	15.138.500
-1000	Arrend. (total)	9	800	86.000	10	266	10.598	9	508.900$	35.292.530.000$	9	507.900	32.471.170.000	10	7.530	8.312.500
	Mista	5	550	94.300	5	133	4.503	5	288.500$	27.264.030.000$	5	291.700	26.680.870.000	4	2.610	2.954.500
1000-	C / própria	10	1.710	418.700	10	493	30.465	10	1.301.400$	267.021.540.000$	10	1.232.700	201.008.050.000	4	3.020	2.832.200
- ..	Arrend. (total)	5	1.250	486.900	5	270	26.762	5	626.200$	107.887.400.000$	5	734.100	148.193.810.000	5	8.730	28.022.900
	Mista	5	810	178.900	6	199	7.095	5	440.800$	48.515.600.000$	5	407.600	40.544.240.000	3	4.720	7.457.000

NOTA: Nesta parte inferior do Quadro os valores calculados para a), c) e d) não abrangem todos os casos (ver Tabela III).

da Tabela III já em termos dos universos estatísticos respectivos, todos confinados à região agrícola do Alto Alentejo.

Conforme pode notar-se, cada estimativa está afectada dum erro padrão — o que lhe confere grande interesse, visto permitir ajuizar do grau de precisão com que foi obtida. Todavia, as estimativas só têm valor prático quando as amostras de que se parta forem seleccionadas por método apropriado.

Numa palavra, não basta apenas atender à precisão das estimativas, é ainda necessário que o método de amostragem garanta a «representatividade» da população em estudo, conferindo-se assim também rigor ou exatidão às estimativas. Apenas o grau das características estatísticas é função do tipo de amostragem e método de avaliação usados.

No caso da «área de trigo» das explorações agrícolas, as seguintes estimativas obtidas com as duas amostras I e II, respectivamente: 117,0 $\pm$ 14,55 e 97,0 $\pm$ 11,09 ha (*Cf.* Tabela III), divergem um tanto, quer quanto aos valores médios, quer quanto aos erros padrões. Mas não podemos afirmar (pelos motivos já estipulados, e por outros ainda a apontar) qual delas tem maior rigor, isto é, qual delas está mais próxima do verdadeiro valor na região agrícola em estudo.

Idênticos considerandos podem fazer-se relativamente às restantes variáveis, bases de estratificação, ou estratos — e daí a razão por que não é agora conveniente levar a efeito porfiados testes de significância, do tipo:

$$t = \frac{(\text{Diferença entre duas médias})}{\sqrt{(\text{Soma dos quadrados dos erros padrões respectivos})}}$$

que, na maioria dos casos, aliás, não oferecem quaisquer dificuldades.

Entretanto, os resultados obtidos para as estimativas, em termos da exploração agrícola, algumas conclusões gerais podem fornecer (apesar dos elevados erros padrões computados), tais como: nítidos acréscimos da «área de trigo» segundo as três classes de grandeza, e acréscimos também evidentes nas respectivas «despesas» e «produções». Outro tanto já não pode dizer-se quando a base de estratificação seja a da forma de exploração, visto serem agora muito menores as diferenças entre as várias cate-

gorias consideradas (apenas o «arrendamento» parece mostrar valores médios mais baixos).

Com as duas restantes variáveis: «pessoal empregado certo» e «gado total da exploração» as conclusões são sensìvelmente as mesmas. Sòmente se regista uma diferença menos nítida entre as médias 798,5 e 1372,5 do número de cabeças de gado das explorações dos estratos de maior grandeza, visto o erro padrão da 2.ª média ser dos mais elevados (com o valor, em percentagem, de: $337{,}00 \times 100/1372{,}5 = 24{,}6\ \%$).

Note-se que as médias *por hectare* para os casos das «despesas com a cultura do trigo» e «produção de trigo» se podem calcular, a partir da mesma amostra de 72 explorações agrícolas, por simples divisão dos valores respectivos da Tabela III, alíneas c) e d), pelos correspondentes das estimativas das médias da alínea a), a da «área de trigo». Os valores que assim se obtiveram na amostra estratificada figuram na Tabela IV.

Eis os valores médios calculados (relativamente às duas bases de estratificação) com a amostra casual: 577$00 ± 65$70 por ha, e 648,50 ± 73,75 kg/ha.

Conforme se indica, pelo método de cálculo usado obtiveram-se, primeiramente, os valores daquelas características «por exploração» agrícola, e só depois se fez a redução ao «hectare». As médias estão, por conseguinte, bem ponderadas, visto ter-se entrado em linha de conta com o factor «área de trigo» das explorações agrícolas; de resto, trata-se simplesmente aqui do uso das fórmulas do método da razão (*Cf.* 2 a 5) em que $\mathbf{r} = S(y)/S(x) = S(w\,r)/S(w)$, com w como variável suplementar (a «área de trigo») e correspondendo r, sucessivamente, a cada uma das variáveis: «despesas por ha» e «produção de trigo por ha».

Quanto aos erros padrões obtiveram-se valores que ficam expressos, em percentagem, em condições análogas aos da Tabela III; outros métodos podem usar-se na sua obtenção, precisando-se apenas do conhecimento prévio da lei da variância dos r relativamente aos w; normalmente, emprega-se a fórmula citada na sub-alínea 2 a 5 para **V (r)**, mas, por vezes, (no caso de ausência de «trend» já referido, etc.) pode fazer-se uso dos valores individuais das próprias razões (YATES 1949 a, p: 197-198 e, ainda p. 212-214).

TABELA III

Estimativas e respectivos erros padrões da « amostra casual » I, e da « amostra estratificada com fracção variável de amostragem » II (Cf. MINISTÉRIO DA AGRICULTURA 1934)

Atributos (explorações c/60 ou mais ha de grandeza)	Tipo de amostragem	1) Grandeza das explorações				2) Formas de exploração			
		Classes de grandeza	N_i	n_i	Estimativas e erros padrões	Estimativas e erros padrões	n_i	N_i	Formas de exploração
a)	I	Geral	1803	72	117,0 ± 14,55 (211.350 ± 26.230)		72	1803	Geral
ÁREA DE TRIGO (em ha)	II	60- 250	716	19	35,5 ± 6,81	117,5 ± 20,10	41	855	CONTA PRÓPRIA
		250-1000	824	33	121,0 ± 21,47	107,5 ± 27,92	21	742	ARRENDAMENTO
		1000- .	263	20	188,5 ± 30,15	136,0 ± 30,54	10	206	MISTA
			1803	72	97,0 ± 11,09	115,5 ± 15,34	72	1803	
b)	I	Geral	1803	81	29,5 ± 2,60 (53.200 ± 4.690)		81	1803	Geral
PESSOAL EMPREGADO CERTO (N.º de empregados)	II	60- 250	716	21	12,5 ± 3,28	30,5 ± 3,03	47	855	CONTA PRÓPRIA
		250-1000	824	39	29,5 ± 2,44	27,0 ± 6,62	23	742	ARRENDAMENTO
		1000- .	263	21	46,0 ± 6,66	30,0 ± 3,68	11	206	MISTA
			1803	81	25,0 ± 1,97	28,5 ± 3,10	81	1803	
c)	I	Geral	1803	72	67.500$00 ± 7.696$50 (121.702.400$00 ± 13.876.700$00)		72	1803	Geral
DESPESAS C/A CULTURA DE TRIGO, em kg (por exploração)	II	60- 250	716	19	21.915$00 ± 4.275$00	70.256$00 ± 11.297$00	41	855	CONTA PRÓPRIA
		250-1000	824	33	62.885$00 ± 8.200$00	59.533$00 ± 12.812$00	21	742	ARRENDAMENTO
		1000- .	263	20	118.420$00 ± 18.645$00	72.930$00 ± 15.457$00	10	206	MISTA
			1803	72	54.720$00 ± 4.932$00	66.149$00 ± 5.813$00	72	1803	
d)	I	Geral	1803	72	75.882,0 ± 8.630,0 (136.815.230 ± 15.559.900)		72	1803	Geral
PRODUÇÃO DE TRIGO em kg (por exploração)	II	60- 250	716	19	23.700,0 ± 4.855,0	82.780,5 ± 12.733,0	41	855	CONTA PRÓPRIA
		250-1000	824	33	79.964,0 ± 13.190,0	65.247,5 ± 14.762,0	21	742	ARRENDAMENTO
		1000- .	263	20	118.720,0 ± 16.195,0	69.930,0 ± 13.918,0	10	206	MISTA
			1803	72	63.274,0 ± 6.756,0	74.097,0 ± 8.712,0	72	1803	
e)	I	Geral	1803	54	786,5 ± 103,90 (1.418.060 ± 187.330)		54	1803	Geral
GADO TOTAL DA EXPLORAÇÃO (n.º de cabeças)	II	60- 250	716	12	170,0 ± 37,10	637,0 ± 89,12	29	855	CONTA PRÓPRIA
		250-1000	824	30	798,5 ± 89,83	925,5 ± 258,70	18	742	ARRENDAMENTO
		1000- .	263	12	1.372,5 ± 337,00	1.047,0 ± 250,80	7	206	MISTA
			1803	54	632,5 ± 65,70	802,5 ± 118,05	54	1803	
« PREÇO DE CUSTO DO TRIGO »		Geral	1803	72	$88,95 ± $06,351		72	1803	Geral

TABELA IV

Estimativas, expressas por hectare, e seus erros padrões duma amostra estratificada (fracção variável de amostragem) de explorações agrícolas da região do Alto Alentejo.

	1) Classes de grandeza		2) Formas de exploração	
DESPESAS DA CULTURA DO TRIGO (Esc. por ha)	60- 250	617$30 ± 120$40	Conta própria	597$90 ± 96$15
	250-1000	519$70 ± 67$75	Arrendamento	553$80 ± 119$20
	1000- ..	628$20 ± 98$90	Mista	536$25 ± 113$65
	Geral	564$10 ± 50$85	Geral	572$70 ± 50$35
PRODUÇÃO DE TRIGO (Kg por ha)	60- 250	667 kg ± 136,75	Conta própria	704 kg ± 108,35
	250-1000	661 kg ± 109,00	Arrendamento	607 kg ± 137,30
	1000- ..	630 kg ± 85,90	Mista	514 kg ± 102,35
	Geral	652 kg ± 59,65	Geral	642 kg ± 75,45

Os valores correctos, assim calculados, diferem dos que possam obter-se pelo método seguido no *Boletim* da Estação Agrária Central (*Cf.* MINISTÉRIO DA AGRICULTURA, 1934), em que as conclusões de maior realce foram tiradas por simples confronto de médias não ponderadas, conforme pode notar-se em todos os seus Quadros sintéticos, a partir do Quadro XX. Ora o factor de ponderação referido tem aqui importância, particularmente nos casos das classes de grandeza superiores — em que é grande a variabilidade entre as explorações e em que as de maior área imprimem efeitos mais exagerados, — e ainda no caso em que é diminuto o número de unidades de tiragem (explorações agrícolas) nos estratos ou na amostra (1).

A incorrecção apontada é um dos «bias» das técnicas de interpretação de resultados, que podem aparecer nos inquéritos por amostragem — e já a isso fizémos referência na Introdução do presente trabalho.

Em relação aos resultados respeitantes ao «hectare», que figuram na Tabela IV, é curioso (embora pouco elucidativo, pelas insuficiências apontadas) destacar o facto das conclusões que deles se podem tirar não coincidirem — pelo menos para a 1.ª base de

(1) Diga-se, em boa verdade, que o relator do trabalho inserto no citado *Boletim* se refere, logo na Introdução, à «insuficiência numérica das respostas obtidas» dando-se conta noutras partes do seu relatório das dificuldades daí derivadas — conforme está bem expresso, por exemplo, na pág. 277, nota inferior.

estratificação — com as anteriormente estipuladas, visto agora se basearem numa mesma área unitária. As diferenças entre as médias dos estratos, tidos em conta os altos erros de amostragem verificados, não atingem agora quaisquer dos níveis de significância, pelo que é apenas lícito afirmar a constância de despesas e de produções.

Se os erros padrões fossem bastante menores, da ordem dos 2 % ou 3 %, ou até mesmo 5 %, certamente muitas das diferenças seriam significativas, o que permitiria dar fundamento aos conceitos económicos correlativos.

Quanto ao importantíssimo problema da determinação do *« preço de custo de 1 kg de trigo »* pouco mais temos a acrescentar ao que ficou referido na sub-alínea 2 a 5 — uma vez tomadas no sentido devido (apenas sob o ponto de vista estatístico, é claro) as considerações que acabamos de expor.

Realmente, em face das razões apresentadas: parca « representatividade » da amostra, « inflator » inapropriado (1), exiguidade de elementos base, altos erros de amostragem, etc., etc., — não devem formular-se quaisquer afirmações definitivas quanto aos estratos e categorias estudados; quando muito, registar-se-á o « preço de custo » geral: $88,95 ± $06,351 (*Cf.* Tabela III).

Este último aspecto dos inquéritos económico-agrários, o da determinação do « preço de custo » dos produtos agrícolas (ou florestais ou pecuários) é sem dúvida de grande complexidade se se tomarem em consideração os múltiplos factores culturais, climáticos, económicos, sociais e mesmo políticos, que os condicionam.

A questão dos afolheamentos e rotações, o tipo de solo, a natureza da cultura principal explorada, a forma de exploração, o tamanho e tipo de empresa, as práticas culturais, etc., e ainda, o grau de eficiência do empresário no uso dos recursos (incluindo os monetários e os de mão de obra) postos à sua disposição, são factores a ter em conta no delineamento dum inquérito agrário por amostragem, de modo que a colheita de dados básicos seja feita de acordo com a finalidade a atingir, embora dentro das condições e circunstâncias impostas.

Diga-se de passagem que, por exemplo, no « National Farm Survey of England and Wales », realizado em 1941/43, muitos

(1) No intervalo entre as duas épocas de recolha de dados, 1930/32 e 1943 houve necessàriamente mudança na distribuição do número de explorações agrícolas pelas 3 classes de grandeza consideradas.

desses problemas foram já focados (*Cf.* OLIVEIRA, 1948) e subsidiàriamente, houve até uma tentativa para definir os denominados «condition indices» das explorações agrícolas inglesas; esta parte do trabalho recaíu, todavia, apenas sobre o material de inquérito de 3 condados e reconheceu-se, por fim, que «the material was not well suited to this kind of treatment...», pois requeria cálculos fastidiosos e um tanto complicados.

Tratemos agora da questão da «representatividade» de universos estatísticos (como os respeitantes às explorações agrícolas de determinado país ou região) a partir de amostras casuais ou estratificadas, etc. Apenas focaremos alguns dos aspectos práticos dessa questão, naquilo em que se relacionem com o caso do inquérito de 1930/32 levado a cabo pela Estação Agrária Central.

Num folheto publicado há poucos anos (DEMING, 1944) o seu autor compendiou os «thirteen different factors that affect the usefulness of surveys» de que vamos dar seguidamente a respectiva lista, por ser deveras de grande interesse geral:

1. Variability in response;
2. Differences between different kinds and degrees of canvass: (a) Mail, telephone, telegraph, direct interview; (b) Intensive *vs.* extensive interviews; (c) Long *vs.* short schedules; (d) Check bloc plan *vs.* response; (e) Correspondence panel and key reporters;
3. Bias and variation arising from the interviewer;
4. Bias of the auspices;
5. Inperfections in the design of the questionnaire and tabulations plans: (a) Lack of clarity in definitions; ambiguity; varying meannings of same word to different groups of people; eliciting an answer liable to misinterpretation; (b) Omitting questions that would be illuminating to the interpretation of other questions; (c) Emotionally toned words; leading questions; limiting response to a pattern; (d) Failing to perceive what tabulations would be most significant; (e) Encouraging nonresponse through formidable appearance;
6. Changes that take place in the universe before tabulations are available;
7. Bias arising from nonresponse (including omissions);

8. Bias arising from late reports;
9. Bias arising from an unrepresentative selection of date for the survey, or of the period covered;
11. Sampling errors and biases;
12. Processing errors (coding, editing, calculating, tabulating, tallying, posting and consolidating);
13. Errors in interpretation; (a) Bias arising from bad curve fitting; wrong weighting; incorrect adjusting; (b) Misunderstanding the questionnaire; failure to take acount of the respondents' difficulties (often through inadequate presentation of data); misunderstanding the method of collection and the nature of the data; (c) Personal bias in interpretation» ([1]).

Conforme se nota, a maior parte das alíneas é de compreensão óbvia, todavia, algumas há que só verdadeiramente serão esclarecidas com a prática. Há pois necessidade imediata de se treinarem equipas de indagadores e técnicos de investigação científica interessados em inquéritos gerais agrários, pois só a experiência ensinará a corrigir muitas das dificuldades e faltas aqui apenas formuladas.

Os programas dos inquéritos precisam de ser prèviamente experimentados e ajustados, quanto aos seus múltiplos aspectos, e, daí, a necessidade das «pilot samples», a que temos feito referência. Tal prática vai sendo já hoje, felizmente, aceite em alguns dos nossos organismos interessados na matéria, como por exemplo, o Instituto Nacional de Estatística e a Estação Agronómica Nacional, e é de desejar que se vá até ao fim — pois há toda a vantagem em realizar duma vez para sempre, e de forma científica, os inquéritos da nossa Agricultura, Comércio e Indústria, e os da determinação da Riqueza e Rendimentos Nacionais.

As perguntas devem ser formuladas, tanto quanto possível em termos quantitativos, mas de nítida compreensão e em conformidade com o nível intelectual das pessoas ou organismos indagados, doutro modo é preferível não as fazer (*Cf.* UNITED NATIONS ORGANIZATION, 1948 b).

([1]) Já depois de escrito o presente trabalho foi-nos gentilmente oferecido pelo referido autor o seu excelente livro «Some Theory of Sampling», ed. John Wiley, 1950, em cujo Capítulo II se faz a revisão da matéria.

Ora no inquérito da Estação Agrária Central há respostas vagas ao quesito «natureza das terras», por exemplo, do género: «fortes na maioria», ou «fortes», ou mesmo «arenosas e argilosas»; teria sido conveniente definir tais termos de modo científico, segundo uma escala de tipos de solos, relacionando as respostas não com a exploração total, antes possivelmente com cada campo de cultura ou «folha» agrícola, etc. Por esse e outros motivos não pudemos fazer a interpretação de todos os resultados do inquérito a explorações agrárias do Alto Alentejo, e muito menos relacioná-los.

O exame crítico das condições do inquérito de 1930-32 à cultura do trigo revela que ele enferma bastante dos erros agrupados nas 13 alíneas da classificação do Professor DEMING. Além dos já apontados acrescentem-se mais os seguintes:

Não havia no início qualquer lista completa de endereços de proprietários agrícolas, de que se pudesse tirar a amostra.

O inquérito demorou cerca de dois anos, tendo os questionários sido enviados em Abril de 1930, só «sucessiva e vagarosamente» chegando as respostas em fins de 1930, durante 1931 e ainda, durante 1932, até Abril, o que nos faz rápidamente crer nos erros catalogados nas alíneas: 1), 2 b), 6), 7), 9) *supra* além de outros.

As respostas foram obtidas, quer pelo correio, quer por indagação directa, tendo-se por fim realizado um «inquérito directo complementar» já depois da última data mencionada (*Cf.* MINISTÉRIO DA AGRICULTURA 1934, pág. II e VII). E muitos outros casos de «erros» se poderia citar, o que não significa que o inquérito seja declarado, sob o ponto de vista estatístico ou outro, de valor negativo ou nulo — pelo contrário, relembre-se o grande papel conferido aos estudos e ensaios prévios, as mencionadas «pilot samples», em qualquer delineamento sério.

Quanto aos «sampling errors» já referidos na Introdução deste trabalho, acrescente-se apenas, com DEMING (*op. cit.*,), que «(they) are often the least of the errors present». Tais erros são inevitáveis, por definição, nos inquéritos por amostragem, e tudo o que há a fazer é procurar uma amostra que conduza a valores pequenos, ou, pelo menos, razoáveis desses erros — o que se consegue por recurso a método de amostragem e grandeza de amostra apropriados. Ora, dum modo geral, no material estatístico agora analisado os erros desta categoria são muito elevados.

Sobre os erros das alíneas 12 e 13 da classificação transcrita

também alguns foram encontrados no próprio relatório que constitui o referido *Boletim* n.º 13 (1934) destacando-se sobremaneira, conforme se disse, os do tipo «wrong weighting» no cálculo das médias das variáveis: «despesas com a cultura do trigo» e «produção de trigo» segundo as classes de grandeza e as formas de exploração (ver especialmente, os Quadros XXVI, XXVIII, XXIX e XXX do *Boletim* de 1934).

O problema da *estratificação* adequada, aquela que permita obter estimativas de grande «precisão» (e de «rigor») em condições de tempo e de dinheiro exequíveis, é dos que deve também merecer ao técnico estatístico estudo cuidadoso. Qual a base de estratificação para um determinado inquérito por amostragem? E como orientar-se, no caso de haver inúmeros atributos qualitativos e variáveis quantitativas, em inquéritos extensos — como os agrários — com seus múltiplos quesitos quanto a áreas, produções e contas de cultura?

Já se sabe que a arte da estratificação consiste em dividir o universo estatístico em classes ou estratos de modo que as respectivas unidades de tiragem de cada um desses grupos sejam tão semelhantes quanto possível, procurando-se desse modo corresponder à finalidade enunciada.

A matéria teórica foi exposta em 1934, com grande desenvolvimento, por NEYMAN, conforme se referiu no nosso anterior trabalho sobre amostragem; recentemente também o citado autor de «On Errors in Surveys» tratou propositadamente dela (DEMING, 1947), suscitando ainda agora grande interesse por toda a parte. Mas são sobretudo os seus aspectos práticos os que nos prendem a atenção neste momento.

Em relação ao caso dum inquérito por amostragem análogo ao da cultura do trigo numa região agrícola ou num país, o que conviria fazer quanto à fase inicial de delineamento? Não dispomos de material estatístico de qualquer inquérito extenso já realizado, conforme se tem dito, e, daí não ser fácil responder cabalmente à pergunta formulada. Mas há já alguma bibliografia mundial respeitante a inquéritos desse género (conforme se verá no capítulo seguinte deste trabalho) que permite dar a conhecer muitos pontos importantes a ter em conta em investigações por amostragem, a fim de obviar a erros como os da lista do Professor DEMING.

Vimos como nos casos simples podem ensaiar-se estatística-

mente algumas bases de estratificação (*Cf.* Tabela III), ansiando-se agora por que o «Plano de Fomento Agrário» nos traga inúmeras séries de dados estatísticos, que poderão permitir outros estudos mais extensos e profundos. Entretanto parece-nos ser de interesse focar, aqui, um dos aspectos do delineamento dos inquéritos complexos por amostragem, o relativo à determinação dos denominados «component indexes».

Este tema foi tratado especialmente no «U. S. Bureau of Agricultural Economics» por MARGARET HAGOOD e outros, tendo sido publicado em 1945 um artigo fundamental, embora breve, sobre a questão, cujo título é só por si elucidativo: «Component Indexes as a Basis for Stratification in Sampling (HAGOOD *et al.*, 1945).

No referido trabalho são apresentados os motivos e a necessidade duma base geral de estratificação quando se trata de inquirir, por amostragem, um conjunto elevado de características nas unidades de amostra; nele também se indicam os métodos de cálculo a usar — através dum exemplo com material do sector agrário, — pelo que se recomenda o seu estudo atento.

A título meramente informativo, indicamos seguidamente as 12 «control characteristics» usadas (pelo método de análise factorial da matriz das respectivas intercorrelações) na construção dos índices necessários à estratificação — em 10 grandes estratos — de 552 «General and Self-sufficing Counties» (HAGOOD, 1945):

«1 Por cent of total population which is rural-farm
2 Altitude measure for County Seat
3 Replacement rate of rural-farm males, 25-69 years of age
4 Per cent change in total civilian population, 1940-43
5 Rural-farm level of living index
6 Per cent change in rural-farm population through migration, 1930-40
7 Mean age of farm operator
8 Per cent farm operators reporting 100 days or more of work off farms comprise of these reporting work off farms
9 Value of home consumption as per cent of values of all farm products
10 Mean value of all farm products
11 Family labor as per cent of all farm labor
12 Value of livestock sold as per cent of value of all farm products».

Depois de cálculos fastidiosos foi determinado um índice para cada um dos 552 municípios, que por fim foram distribuídos por 10 estratos geográficos. E, de modo análogo, se procedeu para com as 6 restantes «major type-of-farming regions» dos Estados Unidos da América do Norte.

A matéria em referência foi assim estudada desde 1943 pelo «Bureau of Agricultural Economics», que chegou depois a estabelecer para todo o território dos Estados Unidos da América do Norte a denominada «The BAE General-Purpose Sample of 101 Counties» (*Cf.* U. S. BUREAU OF AGRICULTURAL ECONOMICS, 1945), não tendo descurado desde então o aprofundamento do problema em vários sentidos económicos e sociais.

Pelo que fica referido já se antevê a enorme tarefa a levar a cabo pelos nossos serviços agrários oficiais. Entretanto, à falta de estudos prévios adequados registe-se que a simples estratificação provincial portuguesa, por exemplo, será já de enorme vantagem nos nossos inquéritos agrários por amostragem, tanto mais que parece ser a estratificação que melhor satisfaz, simultaneamente, os vários organismos interessados na matéria, os Serviços: Agrícolas, Florestais e Pecuários; a Junta de Colonização Interna e, até, o próprio Instituto Nacional de Estatística (pela sua Repartição de Estatística Agrícola).

Dentro de cada província, dever-se-ia ainda procurar obter uma sub-estratificação geográfica, embora leve — pois não convém sub-estratos muito pequenos, que exigiriam fracções de amostragem elevadas. Se a unidade de amostragem for a «exploração agrícola» julgamos que as classes de grandeza ou outros agrupamentos de igual importância deverão ser os estratos imediatos, não se precisando possivelmente de ir mais além.

3. OS ESTUDOS AGRÁRIOS POR AMOSTRAGEM REALIZADOS NO ESTRANGEIRO OU A REALIZAR EM PORTUGAL. DISCUSSÃO GERAL

Seria do máximo interesse dar a conhecer entre nós todos os inquéritos agrários por amostragem já realizados no estrangeiro, pois, desse modo, dispor-se-ia de grande número de informações sobre os mais variados conteúdos e tipos de inquéritos.

Infelizmente não temos à nossa disposição todo esse imenso material; de resto, muito dele não foi publicado — embora seja

acessível em alguns Centros de Investigação estrangeiros. Esperemos que a «Sub-Commission on Statistical Sampling to the Statistical Commission of the U. N. O» possa um dia compilar esses elementos estatísticos e pô-los à disposição dos estudiosos de todos os países.

Ressalvado o motivo das nossas dificuldades, vejamos seguidamente os principais casos estrangeiros de indagações por amostragem cujos relatórios ou trabalhos informativos mais completos se encontram em nosso poder. Uma vez dados a conhecer os seus pontos estatísticos de relevo, dir-se-á então como se devia fazer em Portugal em campos congéneres — o que constitui a matéria da alínea 3 c) com que terminaremos o presente trabalho.

3 a) *Os inquéritos agrários levados a efeito na Inglaterra, Índia e Estados Unidos da América do Norte*

3 a) 1. *O «National Farm Survey of England and Wales» (1942).*

Dos inquéritos realizados em Inglaterra tem particular importância o Inquérito Nacional às Explorações Agrícolas da Inglaterra e País de Gales (*Cf.* MINISTRY OF AGRICULTURE AND FISHERIES, 1946) pelo carácter geral de que se revestiu, quer por abranger todos os condados ingleses, quer pela vasta soma de informações que obteve dos agricultores.

Este inquérito foi realizado durante o período mais sombrio da 2.ª Guerra Mundial, no momento em que se tornava urgente conhecer a realidade da agricultura inglesa, a fim de poder forjar-se um vasto plano de acréscimo de produção. O primeiro estudo, em que as explorações agrícolas eram simplesmente classificadas em 3 categorias gerais, cedo revelou a insuficiência do método seguido na obtenção das respostas de condado para condado, e a pobreza de conteúdo dos questionários.

Pensou-se, então, levar a efeito um inquérito rigoroso, extenso, seguindo um critério único, que por ordem do Ministro seria constituido por 3 partes: a) inquérito directo às explorações agrícolas; b) censo agrícola; c) representação cartográfica, com os limites das explorações (incluindo os dos respectivos campos de cultura), na escala de 1/10.500 ou 1/5.250.

No «Summary Report», publicado em 1946, esclarece-se que a sistematização e análise dos resultados do inquérito fora confiada ao Departamento de Estatística da prestigiosa Estação Experi-

mental de Rothamsted, que se serviu de uma amostra de cerca de 14 % do universo respectivo, o do conjunto das respostas de 290.600 explorações agrícolas com 2 ou mais hectares de grandeza. (As explorações agrícolas ou hortícolas mais pequenas, correspondendo aliás a menos de 1 % da área total, viriam a figurar em inquérito àparte).

Vejamos mais alguns pormenores de carácter estatístico relacionados com este inquérito agrário, de acordo com as informações contidas no relatório o qual inclui, em forma de tabelas, as principais conclusões práticas obtidas, condado a condado, e no conjunto dos condados.

Os agricultores eram visitados por agentes cujo treino prévio constituira «the major problem of the whole survey» devido às condições de guerra; os questionários, uma vez preenchidos, eram enviados para os respectivos «Provincial Agricultural Advisory Centres», em número de 11. Aí eram submetidos a supervisores experimentados e postos então em confronto com os dados do Censo Agrícola de 4 de Junho de 1941, tarefa árdua e longa mas absolutamente necessária, dada a importância do inquérito. Só depois dessa vistoria os dados poderam ser sumariados.

Perante o enorme volume de verbetes preenchidos e a imensidade de variáveis a estudar, e também «on grounds of speed, flexibility and cost», logo se julgou necessário efectuar a interpretação dos dados por amostragem, do que foi encarregado o Departamento de Estatística da Estação Experimental de Rothamsted, conforme se referiu.

Sob a orientação de FRANK YATES (o nosso citado estatístico inglês) e de OSCAR KEMPTHORNE, procedeu-se prèviamente ao estudo experimental dos resultados dos condados de North Riding of York e de Kent, — o que permitiu estabelecer a grandeza e o tipo de amostra que conviria usar para toda a Inglaterra, ao mesmo tempo que dava a indicação da ínfima grandeza dos erros de amostragem a esperar para grandes zonas.

Esta «pilot sample» revestiu-se, deste modo, do maior interesse, tanto mais que se ensaiava pela primeira vez a adaptação do equipamento mecânico à própria análise estatística de dados extensos quando obtidos por amostragem.

A amostra, de cerca de 14 %, ficou constituída como foi referido no nosso anterior trabalho (OLIVEIRA, 1948), isto é, usou-se a «fracção variável de amostragem» conforme os 5 grupos ou

estratos correspondentes às 5 classes de grandeza das explorações agrícolas:

Classes de grandeza (ha):	2-10	10-40	40-120	120-280	280-.
Fracções de amostragem (%):	5 %	10 %	25 %	50 %	100 %

devendo figurar assim na amostra todas as empresas agrárias muito grandes, metade do número das grandes, um quarto do das médias, etc., num total de 39.515.

A tiragem da amostra era tarefa simples, embora fastidiosa, visto haver em cada um dos 11 citados Centros as denominadas «addressograph lists,» já organizadas por condados, com os endereços dos empresários agrícolas, informando juntamente qual a área das respectivas explorações. A amostra estava, pois, automàticamente estratificada por condados, o que lhe conferia, assim, interesse regional.

Resta acrescentar que o «Summary Report» inclui, em apêndice, uma alínea em que se trata largamente do caso do efeito das inevitáveis rejecções e amálgamas, a que pode obviar-se com ajustamentos empíricos apropriados.

Além da pequena informação de carácter metodológico, que aqui se resumiu, o «Summary Report» dedica muitas das suas páginas a esclarecimento das questões respeitantes a:

«Holdings and Occupiers
Tenure, Rent and Length of occupation
Convenience of lay-out and Situation of holdings
Condition of permanent buildings and farm cottages
Type of soil and Natural fertility of soil
Managerial efficiency of occupiers
Water supply and Electricity»

não deixando de tratar também o caso particular dos «condition indices», por nós já referido no fim do capítulo anterior.

Para todos esses temas são fornecidas Tabelas de dupla ou tripla entrada, com os resultados expressos quer em números reais quer em %. Dessa maneira procura-se relacionar as diversas características das explorações agrícolas especialmente com a «grandeza total» (as 5 classes de grandeza já referidas), e o «tipo de exploração» (Pasture, Intermediate, Arable-Types» e «Miscelaneous groups»).

No relatório figuram, também em apêndice, tabelas várias umas

com as produções agrícolas e pecuárias mais importantes, outras com as médias das características para cada um dos condados ingleses. Em qualquer das Tabelas referidas, as percentagens ou os valores quantitativos não são efectados de erros padrões, por estes terem valores pràticamente nulos, conforme se referiu.

3a) 2. *O «Survey of Fertilizer Practice in England»*

Este inquérito começou a ser realizado em 1942 em algumas províncias inglesas, com o fim de se obterem informações quanto ao uso de correctivos e adubações pelos agricultures, procurando-se, ao mesmo tempo, relacioná-las com as verdadeiras necessidades das terras e das culturas.

No delineamento do inquérito colaboraram dois Departamentos da Estação Experimental de Rothamsted, o de Estatística e o da Química. Os trabalhos de campo referentes a colheita de informações e tomas de terra, etc., foram executados sob chefia dos Consultores Químicos do Ministério da Agricultura (YATES, BOYD and MATHISON, 1944).

Cerca de 100 explorações agrícolas (de mais de 2,5 hectares de grandeza) por condado foram incluídas na amostra, como unidades principais; de cada uma delas obtinham-se informações gerais e, ainda, as particularmente importantes para o inquérito restrito sobre as práticas culturais e as adubações e estrumações usadas, quer no ano agrícola em curso, quer nos 4 anos antecedentes.

Para a tiragem da amostra de explorações agrárias usaram-se as «addressograph lists» do Ministério da Agricultura; na amostra, na sua 1.ª etapa, figuravam muito mais «grandes» explorações do que «pequenas», isto é, tratava-se duma amostra estratificada com fracção variável de amostragem. A do condado de Norfolk, por exemplo, tinha a seguinte constituição (YATES, 1949 a):

Classes de grandeza	Fracção de amostragem	N.º de explorações na amostra
Pequenas explorações (2,5 — 20,0 ha)	1/105	22
Explorações médias (20,5 — 121,5 ha)	1/59	36
Grandes explorações (122,0 — . ha)	1/30	9

Na 2.ª etapa de amostragem procedeu-se do seguinte modo: primeiramente, para cada uma das explorações incluídas na 1.ª fase referida formava-se uma lista do número de campos, por cul-

turas (incluindo as pastagens permanentes) e por categorias (duas sòmente: «new-and old-arable lands»); depois tirava-se, à sorte, um campo de cada um dos subgrupos dos dois grupos acabados de referir.

As amostras de terra para análise química eram tomadas casualmente, do conjunto de campos seleccionados em: 1 «old--arable», 1 «new-arable» e 1 «permanent grass-field». Deve dizer-se que nos inquéritos dos anos posteriores passou a tirar-se um campo em cada 3, pois assim já era possível calcular os erros de amostragem desta 2.ª etapa.

Quanto ao inquérito ao agricultor, faziam-se próprìamente perguntas respeitantes à natureza e composição química dos ingredientes empregados em cada ano: estrumações, correctivos cálcicos e adubações.

No trabalho referido apontam-se, descritivamente e através de muitas representações gráficas, os resultados obtidos que, por vezes, acusam «inadequate uses» desses ingredientes. Seria de grande interesse que fossem realizados entre nós inquéritos semelhantes, a fim de nos esclarecermos quanto às necessidades reais da agricultura, agora que a indústria nacional do fabrico de adubos vai constituindo uma certeza consoladora.

Resta acrescentar que os inquéritos desta natureza devem ser delineados de modo a evitar erros do género dos descritos no fim do capítulo anterior. Diga-se de passagem que esses erros são, por vezes, de revelação difícil como no possível caso (apontado por YATES no seu livro sobre amostragem) da tendência de muitos empresários agrícolas em empregarem doses de adubos proporcionais à grandeza dos campos de cultura — facto que obrigaria a que a amostra dos campos de cada exploração fosse tirada com desiguais probabilidades.

3 a) 3. *«The 1942 Census of Woodlands of England and Wales»*

Não temos conhecimento de qualquer publicação completa do relatório respeitante ao Censo Florestal de 1942, pelo que nos reportaremos ao trabalho do Professor DEMING, de 1947 (já por nós referido: OLIVEIRA, 1948), e às indicações de YATES (*op. cit.* 1949 a).

Neste inquérito foi usado o método de amostragem de fases múltiplas. Houve possibilidade de comprovar objectivamente a precisão desse método estatístico, pois obtiveram-se estimativas

através de duas amostras, de 5 %, independentes (mas que se interpenetravam).

A primeira amostra de 5 %, por sua vez, foi constituída por duas amostras de 2 1/2 % cada uma, visto ter havido necessidade de estimativas em curto prazo (meia dúzia de meses), o que só era possível com amostra pequena. Assim o inquérito reduzido revestiu especial relevo, por se poder comprovar a flexibilidade da adaptação dos métodos estatísticos rigorosos às condições experimentais, e por ao mesmo tempo permitir ajustar pessoal e verbas. Note-se que o expediente do uso de várias amostras independentes, por outro lado satisfaz os leigos (a confiança que eles assim adquirem corresponde, aliás, àquela que o técnico estatístico explicita através dos erros padrões apostos às estimativas calculadas).

Com o inquérito pretendia-se avaliar o potencial da Inglaterra em madeiras, e localizar a distribuição das suas matas florestais através de todo o território. Vejamos, agora, como se procedia na aplicação do método de amostragem.

Para ambas as amostras de 5 % (1) tomaram-se como unidades de tiragem, quartos-de-folha das folhas cadastrais de 6 polegadas que eram «systematically located on two interpenetrating 12×10 mile rectangular grids»; todas as manchas florestais com mais de 2 ha de grandeza, e 1 mancha, em cada 5, das áreas menores eram amostradas.

Em cada uma das manchas da 1.ª fase de amostragem demarcavam-se os povoamentos, ficando assim constituídos uns estratos homogéneos sobre os quais viriam a ser localizados os talhões representativos, um apenas por cada povoamento. Era sobre essas parcelas de estudo que se faziam medições apropriadas, em todas ou em parte do número de árvores.

O uso de talhões não casuais poderia introduzir «bias»; tinham-se feito ensaios preliminares, todavia, para adestrar o pessoal técnico, com derrubas totais de alguns povoamentos e medições directas subsequentes para comprovar os cálculos indirectos precedentemente efectuados no mesmo material.

Para concluir, refere-se que a área geral das matas inglesas fora determinada sobre cartas de escala reduzida, — folhas cadastrais de 1 polegada — tendo-se aferido posteriormente, através de

(1) Na 1.ª amostra de 5 %, cada amostra de 2 1/2 % foi efectuada sobre metade de cada uma das «6 inch O. S. quarter sheets».

estudos comparativos com as cartas de maior escala já referidas, a grandeza dos erros nelas cometidos. Para aquela determinação, dentro de cada condado, adoptou-se o processo de amostragem por «linhas».

Acerca de inquéritos por amostragem às matas é ainda de toda a conveniência a leitura do trabalho de FINNEY (1949), e também os de YATES respeitantes a essa matéria.

3a) 4. *«Wireworms and Food Production. A wireworm survey of England and Wales, 1939-1942».*

Neste inquérito a colaboração do técnico estatístico mostrou-se valiosa, pois o problema da determinação da densidade prática de pragas, como a dos *Agriotes*, para toda uma zona ou mesmo para todo um país, oferece sérias dificuldades de amostragem.

Num folheto (MINISTRY OF AGRICULTURE AND FISHERIES, 1944) descrevem-se as várias fases do trabalho e expõem-se com bastante extensão as conclusões práticas obtidas. Quanto à matéria estatística, é-lhe dedicada um capítulo próprio. Para não alongar demasiado o presente trabalho não lhe faremos mais larga referência.

3a) 5. *«Statistical sampling in the Indian Statistical Institute».*

De 1935 até há relativamente pouco tempo, realizaram-se na Índia inquéritos agrários por amostragem delineados quase todos por um conhecido estatístico indiano que fez na Royal Statistical Society uma conferência sobre as actividades do Instituto Indiano de Estatística, cuja importância não podemos deixar de enaltecer (MAHALANOBIS, 1946 a).

As dificuldades de amostragem de territórios como os da Índia são de toda a ordem, diz MAHALANOBIS logo no começo da sua conferência, pois a falta de condições é tremenda; daí resulta que os métodos a usar têm de obviar mais própriamente aos problemas gerais de organização administrativa, do que aos das puras teorias metodológicas. A questão dos gastos monetários e de tempo é a mais importante.

No que se segue apenas apontaremos alguns tipos de inquéritos para a obtenção de estimativas de áreas e produções agrícolas, em Estados como os de Bengala e Bihar; os casos não agrários podem ser estudados no referido artigo de 1946. Vejamos ràpidamente o que se relaciona com o Inquérito às Culturas de Bengala.

O método de «amostragem por áreas» foi usado em 1935 no Estado de Bengala, para a determinação da área e da produção da juta, mas só a partir de 1937 se começou a dispor de material abundante; em 1941 as estimativas eram feitas à escala do Estado. A partir de 1943 o método foi estendido a todas as culturas agrícolas mais importantes.

Por altura de 1945 a província foi considerada subdividida em cerca de 1.100 «cells», cada uma aproximadamente com 16.500 ha; desse modo a «village», a mais pequena unidade administrativa (com 250 ha, pelo menos), ficava nelas totalmente englobada. A amostra compreendia em cada uma das 1.100 «cells», cerca de 100 «grids» ou unidades de amostra; para a tiragem dentro de cada uma das «cells», seguia-se o método dos pontos casuais, a cada «grid» correspondendo um ponto.

As unidades de tiragem tinham à volta de 1 ha; uma vez marcadas nos mapas cadastrais as unidades de tiragem que formavam a amostra, fazia-se o inquérito a todos os campos nelas incluídos (os campos indianos têm, em média, áreas de 1/4 ha a 2 1/2 ha), todavia desses campos apenas se ceifavam pequenos talhões (com 1 a 10 m^2) da cultura em estudo.

Conforme se nota, a fracção de amostragem é de cerca de 1/800, todavia suficiente para permitir a obtenção de estimativas com erros de amostragem muitas vezes inferiores a 2 %.

Ainda a propósito do Inquérito às Culturas de Bengala esclarece-se que o autor usava sempre duas amostras que se interpenetravam, o que lhe permitia assegurar-se do valor prático do método usado, e quase sempre obviar aos inconvenientes dos «biased errors». Outra particularidade dos inquéritos indianos é a do uso de aparelhagem portátil para ceifa e pesagem da porção de cultura agrícola incluída nos pequenos talhões de estudo, em cada campo amostrado.

O artigo que vimos citando refere ainda outros casos de interesse mais pròpriamente social, embora ainda sobre as populações rurais. Entre eles destaca-se o «Sample survey of the economic background of the Bengal famine», que foi realizado em 1944-45 sobre 15.769 famílias «selected at random from 386 villages, which themselves were selected at random from 41 (out of 86 rural) sub-divisions (administrative units) covering about 60.000 sq-miles in Bengal» (MAHALANOBIS 1946 a, p. 366).

O inquérito subsequente, à fome de Bengala, permitiu obter

estimativas com erros de amostragem de 3 a 5 %, e foi efectuado em curto prazo. No « Combined agricultural, economic, and population survey in Bengal, 1945-46 » o número de famílias rurais inquiridas ultrapassou 80.000.

Noutro artigo (MAHALANOBIS, 1946 b), respeitante ao Estado de Bihar, são dadas informações pormenorizadas quanto aos aspectos teóricos e práticos dos respectivos problemas de amostragem; como apresenta poucas alterações básicas, salvo melhor sistematização das matérias, não lhe daremos aqui maior relevo.

Um outro conjunto de inquéritos com interesse agronómico vem sendo efectuado pelo « Indian Council of Agricultural Research ».

3 a) 6. *Os inquéritos norte-americanos por amostragem, no Estado de Iowa e em parte dos Estados da Flórida e da Califórnia.*

Vejamos com pormenor, (de acordo com o artigo de JESSEN já citado, OLIVEIRA 1948), o caso dos inquéritos agrários do Estado de Iowa, correspondentes aos anos de 1938 e 1939. Num importante trabalho (JESSEN, 1942) os resultados obtidos em cada ano são primeiramente analisados em separado e, depois, em conjunto. A denominada « matching sample », correspondente a parte dos dados obtidos nos dois anos consecutivos, é também focada.

O inquérito de alguns dos distritos dos Estados da Flórida e da Califórnia vem relatado num trabalho posterior (JESSEN *et al* 1944), e a ele se fará também referência, por tratar de casos agrários de grande interesse.

O folheto de 1942 revestiu-se de importância particular, quer devido à soma de factos indagados, quer pelas valiosas indicações estatísticas formuladas, — inclusivé, as referentes ao problema da eficiência dos diversos métodos de amostragem em termos dos gastos monetários nas várias fases de execução dos inquéritos.

Na realização do inquérito a todo o Estado de Iowa cooperaram 3 organismos: a « Iowa Agricultural Experiment Station » (pela sua Secção de Estatística), o « Bureau of Agricultural Economics » e o « Agricultural Marketing Service ». Eis como se procedeu na amostragem de 1938 (a do ano seguinte cingiu-se aos mesmos moldes), na qual se fez o inquérito directo a cada exploração agrícola da amostra, com preenchimento de questionário apropriado.

Cada um dos 101 municípios do Estado (um desses municípios fora prèviamente subdividido em dois) constituia um estrato. Como unidades de amostragem figuraram os quartos-de-secção (1), em número de 908, ou seja cerca de 0,42 por cento do total de 219.176 dessas secções em que se considera sub-dividido o Estado de Iowa; todas as explorações agrícolas (de mais de 1,2 ha de área) dentro das secções incluídas na amostra eram inquiridas.

Para descrever o processo de tiragem da amostra vamos empregar as próprias palavras de JESSEN (1942). «Townships were selected at random from each country — seven from an average size county of 16 townships, and in the some proportion for counties of other sizes. Quarter-sections were selected at random from each for the selected townships — one each from five, two each from the remaining two. An average size county would therefore have nine quarter-sections selected for the sample».

É interessante fazer agora o confronto entre os dados referentes a todo o Estado de Iowa e os da amostra do inquérito de 1938:

Municípios		Freguesias	Quartos-de-secção	Número de explorações
«Universo» . . .	101	1600	219.176	213.318
«Amostra» . . .	101	700	908	773

Conforme pode notar-se, apenas 773 explorações agrícolas foram inquiridas — e todavia permitiram tirar conclusões de grande precisão (erros de amostragem da ordem dos 3 a 4 % para a maior parte dos atributos respeitantes a áreas, produções, gados, valor dos produtos, despesas e receitas).

É certo que o Estado de Iowa, que tem quase o dobro da área de Portugal metropolitano, é tido como de grande homogeneidade; apesar disso, entretanto, há que encarecer o valor científico do método de amostragem usado e o grau do seu ajustamento à realidade agrária inquirida.

Antes de referir outros aspectos estatísticos de interesse esclarece-se que os inquiridores estavam frequentemente em contacto com a Secção de Estatística da Estação Agronómica de Iowa, a fim de atender ràpidamente a qualquer dificuldade surgida. Cada inquiridor gastou, em média, 32 minutos a preencher cada um dos questionários. Para a localização da exploração agrícola usaram-se

(1) Cada «quarter-section grid» tem cerca de 65 ha.

mapas adequados ou fotografias aéreas, havendo ainda possibilidade de recurso aos esclarecidos « county AAA officers ».

O inquérito por amostragem, de 1939, incidiu sobre 897 quartos-de-secção, dos quais cerca de 50 % correspondiam aos do inquérito de 1938: por esse meio descortinaram-se muitas e interessantes possibilidades, ao mesmo tempo que se ampliaram os conhecimentos estatísticos sobre a não fácil matéria dos inquéritos agrários por amostragem.

No seu folheto de 1942 JESSEN tratou muitos dos aspectos respeitantes à já citada « matching sample », e tirou conclusões, que provaram ser de grande valor, sobre o célebre delineamento da « Master Sample of Agriculture » (a referir noutra sub-alínea).

A propósito do inquérito das mesmas explorações em mais de um ano, o autor chama a atenção para a categoria de erros denominados « erros de memória » mostrando que os agricultores, e, dum modo geral, todas as pessoas objecto de inquérito, estão sujeitas a eles, o que deve exigir cuidado especial da parte do indagador, pois nalguns casos a pergunta formulada em outra época é já de sentido diferente, ou interpretada como tal.

Perante o material colhido nos dois anos, o autor estudou também as eficiências relativas de diversas grandezas da unidade de amostra (para 1/4, 1/2, 2, 4, 36-secções, e para a empresa agrícola), e o efeito do uso de diferentes bases de estratificação (geográficas: por freguesias, municípios, regiões agrícolas; ou por « classes » e « formas » de exploração). Pôde então concluir que, em geral, as secções menores são mais eficientes e que a estratificação geográfica apresenta também vantagens, mas os estratos das formas de exploração não trazem acréscimo da eficiência, etc.

Quanto ao método de avaliação de estimativas, o autor discute 3 modalidades do « método da razão » consoante a variável suplementar corresponde a: número de secções, grandeza da exploração ou número de explorações. A modalidade referida em 2.º lugar mostrou-se mais eficiente.

Muitas outras questões foram tratadas, destacando-se, sobremaneira, a referente aos gastos monetários no capítulo intitulado « The problem of maximising amount of information obtainable from a given expenditure by varying size of the sampling unit and the number taken », em que são estabelecidas fórmulas para aplicar, na prática, em inquéritos similares.

Em apêndice do trabalho referido são cotejados criticamente, um a um, os resultados obtidos em termos do universo (por «expansão» da amostra) com os correspondentes aos dos organismos superiores da estatística oficial: do Censo Federal, do Censo Estadual, e do «Agricultural Marketing Service» (A. M. S.). As discrepâncias por vezes verificadas são atribuídas pròpriamente aos vários critérios usados e às diferenças de conteúdo das perguntas formuladas, — de mais rigor científico no caso da amostragem.

Daí a justificação de conclusões tais como (*op. cit*, p. 94):

«... One of the most important conclusions to be drawn from these comparisons is that none of the agencies can be said to provide absolutely accurate information».

«... The sample survey has agreed quite well with the other agencies in greater number of items. Some of the large discrepancies have been explained as due ... to questionnaire differences ...»

«... In many cases it appears that the sample survey was more accurate than the A. M. S., especially with those items appearing on a great number of farms. Furthermore it seems reasonable to believe that in some items the sample survey was more accurate than the census».

«... In general the sample survey has proved to be satisfactorily representative. With the exception of a few items of usual minor importance, errors attributable to sampling have been reasonably small».

Resta-nos tratar dos casos dos inquéritos dos Estados da Flórida e da Califórnia, cujas estruturas agrárias são mais heterogéneas do que a do Estado de Iowa (JESSEN *et al*, 1944). O caso da Flórida é bastante análogo ao da Califórnia; como o folheto de 1944 desenvolve mais este último, será dele, portanto, que nos ocuparemos.

Uma das questões que suscitou dificuldades, foi a relacionada com o facto de muitas das explorações agrícolas serem constituídas por vários tratos não contíguos de terreno, tal como acontece, aliás, em Portugal (só no concelho de Colares, por exemplo, a cada proprietário correspondiam 4,5 prédios; *Cf.* BARROS 1938).

Daí o ter de definir-se mais pormenorizadamente o que deve entender-se por exploração agrícola «interior» às unidades de

amostragem, agora constituídas por meias-secções (cada uma com cerca de 130 ha, em média).

Quanto ao indicador, que define a carácter «interior» da exploração agrícola, adoptara-se o do assento de lavoura respectivo e, à falta deste, o do seu «ponto Norte»; no caso da exploração agrária ser contituída por várias courelas usar-se-ia o «ponto Norte» da courela ou prédio mais ao Norte, — mas ainda dentro de unidade de amostragem demarcada no mapa das estradas municipais ou sobre as fotografias aéreas.

Acentue-se que o uso da meia-secção como área de amostragem era aqui uma necessidade, pois não se conhecia o «número de explorações» para servir de variável suplementar; por outro lado, essas secções não precisavam de ter exactamente a mesma área, antes deviam corresponder a limites naturais ou de fácil distinção.

Conforme se nota, cada exploração agrícola, constituindo neste caso a «unidade de observação», é associada com uma, e só uma, meia-secção, pelo que qualquer dessas unidades tem a mesma probabilidade de ficar incluída na amostra, — que no caso da Califórnia correspondeu a 5, 6 % do número total de meias-secções. Desse modo, a amostra é isenta de tendenciosidade («bias»).

O último aspecto que vale a pena deixar aqui citado diz respeito às diversas bases de estratificação ensaiadas pelos autores: a das freguesias, municípios e regiões agrárias; a do «número esperado de explorações», com 5 classes:

I. 0; II. 1-3; III. 4-6; IV. 7-11; V. 12 — ou mais construções agrícolas apresensadas nos mapas; e ainda, a do «type-of-farm»:

1. «Stock-ranches»; 2. «Grain farms»; 3. «Tree fruit and nut farms»; 4. «General farms»; 5. «Dairy farms»; 6. «Poultry farms»; 7. «Small-scale farms».

Qualquer destes estudos foi possível devido à existência de bons mapas e de fotografias aéreas postos à disposição dos orientadores do inquérito. Por isso, as conclusões tiradas têm valor científico e devem merecer a atenção dos estudiosos dos problemas da amostragem.

3a) 7. *«The United States Master Sample of Agriculture, 1945».*

O estabelecimento da «amostra principal» das explorações agrícolas norte-americanas coube ao Laboratório Estatístico da prestigiosa Estação Agronómica de Iowa, em íntima colaboração com o «Bureau of Agricultural Economics».

Os trabalhos preliminares começaram em 1943. Tratava-se de obter uma amostra de explorações agrícolas que permitisse, uma vez estabelecida (mas periòdicamente reorganizada), o estudo de grande número de factores agrários, e mesmo económicos e sociais respeitantes a todos os Estados da grande União Norte-Americana. Depois de estudos vários foi resolvido usar o método de amostragem na base de áreas, para se poder obter representação geográfica adequada.

De acordo com o artigo publicado no volume 40 do Jornal da Associação Americana de Estatística, em fins de 1944 (KING and JESSEN, 1945) a amostra principal incluia 67.000 «áreas», que englobavam um conjunto de cerca de 300 mil explorações agrárias, ou seja, aproximadamente 1/18 da área da União, correspondente a 1/18 do número de explorações agrícolas. Todos os 3.070 Municípios da União figuravam na amostra.

Como grande estratificação geral estabeleceram-se 3 estratos «... on the basis of incorporation and density of population. The total area of the United States was classified into: (i) incorporated areas, (ii) unincorporated areas relatively densely populated, (iii) unincorporated areas sparsely populated» (*op. cit.* p. 47). Acrescente-se que o número total de explorações ficou distribuído pelos 3 estratos respectivos, nas percentagens de: 3,4 % 5,9 % a 90,7 %.

Pondo de parte muitos dos pormenores respeitantes à marcação dos estratos, na qual se recorreu largamente às fotografias aéreas já existentes para quase todo o território, vejamos alguns aspectos da tiragem das unidades de amostra, aqui constituídas por pequenas áreas, cada uma com 650 ha em média (sem que tivessem de ser, necessàriamente, da mesma área, pois chegavam a variar desde 180 ha, Estado de Indiana, a 28.000 ha, Estado da Nevada).

As unidades de amostra incluiam geralmente 4, 5 ou 6 explorações agrícolas. No terceiro grande estrato referido, também denominado «open-country», procedeu-se, em primeiro lugar, à demarcação, nos Mapas das Estradas principais, de outras unidades maiores, as «count units» correspondentes a limites administrativos ou a limites naturais, e de modo a incluir de 6 a 30 explorações. Cada uma dessas unidades desse estrato englobava, pois, uma ou mais unidades de amostra constituídas pelas pequenas áreas já referidas.

Depois da formação de totais de vária ordem, que por sua vez eram seriados e inscritos nos mapas de cada município, procedia-se à tiragem da amostra de «count units»; essa amostra englobaria apenas as referidas unidades que incluissem cada 18.ª unidade de tiragem mais pequena (para número inicial tomava-se casualmente um número entre 1 a 18). Finalmente, de cada «count unit» seleccionada tirava-se, também de modo casual, o número requerido de unidades de amostragem menores.

Conforme se nota, pelo método indicado apenas eram subdivididas as «count units» que figurassem na amostra, o que poupava imenso trabalho e tempo. Depois de delimitadas nos mapas as unidades menores, o inquérito far-se-ia apenas às respectivas explorações agrìcolas nelas incluídas.

Quanto aos outros dois grandes estratos, foram também gizados métodos próprios, sem que se necessitasse das «count units» intermédias; para o caso de estrato (i), que englobava dois grupos principais de «incorporated places» dentro de cada município — os de menos de 2.500 habitantes e os de 2.500 ou mais habitantes — os autores do plano recorreram-se dos dados rigorosos do «Bureau of the Census», quanto ao número e localização das respectivas explorações agrícolas, etc., o que facilitou a tarefa. (É interessante referir que essa repartição americana preparava paralelamente a «Master Sample of City Areas»).

Realce-se mais uma vez o facto daquela amostra permitir os mais variados estudos, muitos deles sobre sub-amostras adequadas. Bastaria a possibilidade de se obterem estimativas rigorosas na previsão de colheitas, por exemplo, para justificar plenamente que se estabeleça entre nós uma amostra semelhante. Acresce ainda que, dado o emprego dum método como o de «amostragem por areas», as mudanças que se verifiquem em pequenos períodos de tempo não se reflectem grandemente na amostra; de resto, são possíveis (e sob muitos aspectos necessárias) reorganizações periódicas.

3 b) *Outros pontos àcerca dos inquéritos por amostragem ainda exemplificados com o caso do inquérito à cultura do trigo do Alto Alentejo*

Depois dos diversos relatos das sub-alíneas anteriores já podemos tratar de mais alguns pontos respeitantes ao inquérito cujos

elementos base foram objecto de análise estatística no Capítulo II deste trabalho.

No uso que então se fez dos resultados das respostas ao inquérito da cultura do trigo, realizado em 1930-32 pela Estação Agrária Central, foram-se apresentando os principais processos de amostragem — pelo menos sob os seus aspectos práticos gerais. Depois dessa exemplificação, tentou-se a análise de muitas das características dessas explorações agrícolas do Alto Alentejo, tendo-se estabelecido algumas conclusões. Mas como os dados estatísticos eram bastante restritos e incompletos, o «inflactor» era inadequado e a amostra não pôde ser considerada como representativa da região inquirida obtiveram-se, assim, conclusões precárias quanto ao valor prático.

Ainda o facto de apenas se poder expandir a amostra de explorações de 60 ou mais hectares de grandeza obsta a que se possam comparar todos os valores das estimativas do universo estatístico, o Alto Alentejo — obtidas a partir daquela amostra — com os da Estatística oficial. Seria interessante fazer esse confronto para cada uma das características: área total, área de trigo, produção obtida e gado existente, pois poderia revelar, ou pontos de concordância, o que de qualquer modo seria um tanto comprovativo, uode divergência, o que poderia sugerir aos nossos Organismos estatísticos interessados modificações no processo de recolha de dados.

A estimativa, à escala regional, que pode fàcilmente obter-se a partir da amostra casual (mas já não se consegue calcular o respectivo erro padrão) é a da área de trigo. Assim, tomando em linha de conta todos os valores daquela variável nas 122 explorações agrícolas inquiridas, por emprego do método da razão (a variável suplementar x correspondendo aqui à área total, e **X** tendo o valor 1.343.833 ha — *Cf.* JUNTA DE COLONIZAÇÃO INTERNA, 1943) tem-se:

$$\mathbf{Y} = \frac{12.236}{85.208} \times 1.343.833 = 192.970 \text{ ha}$$

Ora, a área «semeada» de trigo em 1929 foi de 163.743 ha, e, em 1930, de 169.263 ha, enquanto que em 1931 acusa o valor 182.312 ha (para 1932 o número indicado no «Boletim de Informação e Estatística Agrícola» é 220.924 ha); no quinqénio 1927-

-1931 a média é de 168.830 ha, subindo para 181.050 ha no quinquénio 1928-1932.

Em qualquer dos casos a estimativa calculada por «expansão» da amostra é mais elevada, com uma diferença de 6,6 % em relação à média do quinquénio 1928-1932, que atinge 14,3 % se a compararmos com a média do 1.º quinquénio considerado, o que é mais lógico. (Mesmo que a aproximação fosse flagrante não ficariam invalidadas, contudo, muitas das considerações que fizemos sobre os possíveis «biased errors» da amostra).

Para o caso da variável «produção de trigo» o total da colheita a partir da amostra de apenas 110 explorações agrícolas (o conjunto das 122 explorações inquiridas com exclusão das de número de ordem 1, 13, 14, 26, 31, 38, 50, 91, 94, 96, 117, 123) é:

$$Y = \frac{7.181.700}{75.395} \times 1.343.833 = 128.606.850$$

enquanto que os valores da Estatística oficial (supondo que o trigo tinha um peso específico médio de 78 — número este indicado em respostas de alguns empresários agrícolas) eram: 98.901.500 kg para 1929, 120.960.200 kg para 1930, 117.201.500 kg para 1931 e 211.308.800 kg para o grande ano agrícola de 1932, — com a média 101.719.300 kg no quinquénio 1927-1931 e 123.123.600 kg para 1928-1932.

Conforme pode notar-se, a estimativa agora obtida apresenta um valor também muito distanciado da média de produção do 1.º quinquénio.

Note-se que, possìvelmente muitas das respostas ao inquérito vinham já influenciadas pela euforia económica resultante principalmente da Campanha do Trigo com seus «prémios de cultura e de arroteia» (*Cf.* MINISTÉRIO DA AGRICULTURA 1934). Outras justificações poderiam apresentar-se, mas não pretendemos ir agora mais além nesse sentido. Diga-se, de passagem, que a falta de rigor das informações da lavoura no preenchimento dos manifestos é hoje ponto incontroverso, e daí a necessidade de se modificar o método de colheita de dados.

E, de modo análogo se procederia no que respeita às outras características; simplesmente encontram-se, logo, maiores dificuldades, qualquer que seja o caso considerado. Com a amostra estratificada, seria fácil mostrar que o método da razão exagera a

estimativa da área de trigo, pois o valor obtido para a classe de grandeza inferior, a das explorações agrícolas com menos de 60 ha, é muito elevado; trata-se, possìvelmente, dum estrato especial (nas explorações inquiridas a percentagem da área de trigo de cada uma delas, em relação à área total respectiva, é muito alta), e o referido método de avaliação de estimativas parece ser aqui inadequado.

A dificuldade, que acaba de indicar-se, deveria ser motivo de estudo crítico aprofundado — todavia, as muitas limitações apontadas nada mais permitem adiantar nesse sentido; o caso do «preço de custo do trigo», que seria de tanta importância analisar, também enferma desses mesmos defeitos. Se as pequenas explorações agrárias alentejanas tivessem sido objecto do inquérito da Junta de Colonização Interna, — pelo menos por amostragem — outras seriam, sem dúvida, as possibilidades de confronto das estimativas obtidas com os dados da Estatística oficial.

3c) *Contribuição para o delineamento de inquéritos agrários por amostragem a realizar em Portugal e no Ultramar.*

Através de todas as páginas anteriores viu-se como é possível efectivar o estudo de numerosíssimos problemas nacionais por meio de amostras representativas, quando estatìsticamente delineadas em concordância com a realidade dos domínios ou sectores de pesquiza.

Fundamentalmente, sobressai de todas essas páginas a seguinte conclusão prática; — é possível estudar, por processo rápido, eficiente (e com muito menos dispêndio monetário do que era necessário noutros tempos) qualquer sector ou região importante, desde que se recorra às técnicas de amostragem.

É mesmo possível levar tal indagação científica a campos difíceis ou a regiões atrazadas, — o caso das nossas grandes províncias ultramarinas parece impôr-se-nos como tarefa gigantesca de inquérito, antes de qualquer planificação progressiva de aproveitamento geral das suas riquezss.

Antes, porém, entendemos que, como tarefa imediata, se deve começar pela Metrópole portuguesa esse há tanto tempo necessário inquérito científico como condição de porfiados estudos ulteriores. Vejamos, portanto, o caso de Portugal, mais particularmente o do seu sector agrário. Claro que se trata duma mera contribuição

para o esclarecimento do problema, tanto mais que apenas são focados alguns dos seus aspectos estatísticos (1).

Supunhamos que se dispõe de fotografias aéreas de todo o território, (não havendo, porém, os correspondentes mapas cartográficos) e que se podem usar as antigas folhas da «Carta Agrícola» e as modernas da «Carta do Estado Maior do Exército». Um *delineamento* simples e exequível para Portugal metropolitano, nas presentes condições, comportaria, em termos gerais:

Amostra estratificada, com fracção uniforme de amostragem da ordem dos 10 % da área, para obtenção de estimativas à escala regional, mas que pode baixar para 5 % se os estratos forem os correspondentes às áreas das nossas províncias — ou mesmo 1 %, se apenas se requerem estimativas para o conjunto do país.

Como unidades de amostra, embora não tenhamos à nossa disposição elementos portugueses de qualquer inquérito similar, quer parecer-nos, entretanto, que se deveriam usar unidades com áreas de 50 ou 100 ha (a ultima mais particularmente indicada para a grande região alentejana, a de 50 ha para o resto do país).

A tiragem da amostra de 5 %, com um total de 8850 unidades de amostra (na hipótese destas terem todas apenas 50 ha de área), far-se-ia casualmente dentro de cada grande estrato referido; isto é, cada estrato teria uma representação proporcional à respectiva área.

Quanto à questão das estimativas das áreas das culturas importantes, em cada unidade de amostra seriam marcadas todas as parcelas culturais de grandeza compatível, e as respectivas áreas medidas com planímetro.

---

(1) Devemos esclarecer que para a definição de muitos dos pontos do delineamento aqui proposto houve, por várias vezes, troca de impressões com alguns dos nossos técnicos estatísticos do Instituto Nacional de Estatística, em particular com os Drs. PAES MORAES e G. CASTRO E SILVA e com os Eng.os Agrón. V. PINTO DE MAGALHÃES e F. FARIA VAZ.

Também travámos discussão útil sobre esta matéria com o Eng. Agrón. L. SOUSA E FARO (D. G. S. A.) e com alguns dos membros da Comissão Orientadora do «Plano de Fomento Agrário», especialmente com os Eng.os Agrón. J. Q. PESSOA LOPES (D. G. S. A.), MÁRIO PEREIRA (J. C. I.) e A. CÂNDIDO FERREIRA (J. C. I.) e os Eng.os Silv. MÁRIO A. DOS SANTOS GALLO (D. G. S. F. A.) e JOSÉ MARIA DE CARVALHO (D. G. S. F. A.).

Quanto ao problema das produções, armentio e controle das características económicas das explorações agrícolas, entendemos ser necessário fazer uma sub-amostragem posterior dentro de cada unidade de amostra principal — o que teria de ser realizado depois do levantamento das áreas culturais, numa 2.ª fase de execução do inquérito por amostragem.

Fica assim esboçado, a traços largos, um esquema que poderá usar-se nos inquéritos agrários da Metrópole. Muitos pormenores haveria ainda a esclarecer, como o do critério a seguir no caso da unidade de amostra cortar o estrato, quais as culturas estremes ou não estremes a inquirir, os respectivos critérios de definição, etc., etc. — mas isso constitui pròpriamente matéria da comissão técnica que for nomeada para orientar os trabalhos.

Entretanto, deve-se ainda salientar que o valor de cada inquérito depende muitíssimo do treino prévio das equipas de campo, as quais devem ser industriadas, em conjunto, quanto aos métodos, normas e critérios gerais a usar na colheita dos dados. Ainda dentro deste espírito, chama-se a atenção para a necessidade de realizar ràpidamente todas as operações de campo, uma vez iniciados os trabalhos do inquérito agrário.

Com o delineamento proposto supomos que para as estimativas a obter, pelo menos naquelas características mais frequentes em cada estrato, os erros de amostragem resultantes seriam pequenos. As culturas extremamente localizados deverão ser estudadas em inquérito especial.

Agora que os estudos para o «Plano de Fomento Agrário» estão em curso é da máxima importância obter a maior quantidade possível de factos agrários, pelo que as técnicas modernas de amostragem deviam ser afoitamente usadas (1); conforme se referiu no 1.º Capítulo, seria também conveniente estabelecer conjuntamente um esquema geral de experimentação.

Quanto aos inquéritos do *Ultramar português* julgamos ser absolutamente necessário procurar realizá-los no mais curto espaço de tempo, pelo menos logo a seguir à conclusão dos inquéritos da Metrópole; nenhum plano nacional de aproveitamento das nossas

(1) Depois de escrito este trabalho tivemos conhecimento que a Comissão Orientadora do «Plano de Fomento Agrário» resolvera empregar essas técnicas no seu inquérito, segundo delineamente similar ao aqui indicado.

riquezas pode ser considerado eficaz se se desconhecerem as condições e as possibilidades dos territórios ultramarinos, em especial os de Angola e Moçambique (*Cf.*, por exemplo, CORREIA 1950).

Claro que as dificuldades de realização desses gigantescos projectos são agora maiores, por grande carência de elementos básicos para qualquer inquérito: mapas cadastrais, cartas agrícolas (e florestais), fotografias aéreas, etc. As possibilidades de transporte e acesso às zonas interiores são também muito limitadas.

Entretanto, algo é possível fazer; de resto, sabe-se que estudos similares têm sido realizados em regiões extensas e atrazadas, (MAHALANOBIS 1946, SUKHATME 1949) e não devemos esquecer a flexibilidade dos métodos de amostragem, sem por esse facto deixarem de ser rigorosos e expeditos. Eis os principais pontos a ter em consideração.

Primeiramente, sem dúvida, haveria que formular as finalidades dos inquéritos e definir quais as condições presentes dos territórios ultramarinos. Às comissões que viessem a formar-se, e que incluiriam necessàriamente pessoal altamente conhecedor dessas regiões, haveria que agregar técnicos estatísticos especializados.

Em seguida far-se-ia um delineamento, digamos, experimental, isto é, do tipo «pilot sample» — cujo valor já tantas vezes foi aqui enaltecido. Só depois desse estudo prévio se poderia pensar no delineamento rigoroso, a executar, então, segundo normas ajustadas às realidades africanas.

Sem se querer avançar muito mais na matéria (por se desconhecerem quase totalmente os problemas coloniais) não podemos, entretanto, deixar de emitir a opinião de que os estratos devem ser bastante extensos, obedecendo fundamentalmente às grandes divisões naturais — o que permitirá usar uma baixa fracção de amostragem (de 1 %, para fixar ideias), única alternativa de se realizar o inquérito agrário, completamente e em pouco tempo: as duas premissas das responsabilidades públicas dos orientadores de qualquer inquérito geral desse tipo.

Ao formular-se aqui o problema da amostragem, parte-se da hipótese de que, para aqueles grandes estratos apenas se pretendem estimativas das principais áreas cultivadas e arborizadas, embora com «precisão» razoável. Mesmo que a fracção de amostragem se

afaste um tanto da proposta, de qualquer modo julgamos que o método de amostragem na base de «áreas» está indicado.

Qualquer das modalidades de amostragem por etapas múltiplas e por fases múltiplas (ou ambas) deverá ser então seguida, visto permitirem grandes facilidades quanto ao importante aspecto da organização dos inquéritos, conforme anteriormente se referiu.

Quanto aos territórios ultramarinos de mais pequena extensão do que os de Angola e Moçambique, desde que existam elementos cartográficos satisfatórios torna-se fácil estabelecer os métodos de amostragem apropriados.

Por tudo o que acabamos de expor, quanto ao actual desconhecimento da verdadeira natureza do material estatístico do Ultramar português, entendemos que os Organismos competentes deveriam ir desde já adestrando pessoal, para que um dia pudessem cumprir cabalmente a sua patriótica missão.

4. SUMMARY

STUDIES ON AGRONOMIC STATISTICS. V. — THE MODERN STATISTICAL SAMPLING SURVEYS FOR STUDYING THE PORTUGUESE AGRARIAN PROBLEMS.

1. The practical value of sampling in agricultural surveys is stressed. The complementary character of surveys and experimental trials is strongly emphasized.

2. By using data of a judgment sampling survey about the cultural conditions of wheat carried out in 1930/1932 in the region of «Alto Alentejo» many modern sampling methods are described.

Some statistical interpretation of the data is given; high sampling errors have been computed. A critical explanation of the sources of the wheat survey has shown that biased errors possibly arose.

3. A brief statement of representativeness of the wheat survey is given. Attention is required fo the component-indexes as a basis for stratification in the modern sampling surveys.

Finally a general sampling design for the agricultural surveys in Portugal is suggested.

REFERÊNCIAS BIBLIOGRÁFICAS

BARROS, H. DE

1938 A região de Colares. *Ser. Estudos Inf. Técnica, Minist. Economia*, Lisboa, **5**.

BERNAL, J. D.

1943 *The Social Function of Science*, London (4th Édition).

CORREIA, ARAÚJO

1950 O programa económico nacional. *Diário das Sessões* (15 de Abril de 1950).

DEMING, W. E.

1944 On errors in surveys. *Amer. Social Rev.* **9**: 359-369.

1947 Allocation in stratified samples. (25 March 1947, a ciclostilo).

DIAS JR., J. N. FERREIRA

1946 *Linha de Rumo. Notas de Economia Portuguesa*, 1.º volume. Lisboa.

EISENHART, Ch.

1948 The role of a statistical consultant in a research organization (presented at the 25th Session Intern. Stat. Institute, Washington, D. C.).

FINNEY, D. J.

1949 Random and systematic sampling in timber surveys. *Forestry* **22**: 64-99.

GUERREIRO, A.

1948 Amostra estatística. Métodos usados na sua escolha. *Rev. Economia*, Lisboa **1** (4): 221-224.

HAGOOD, M. J. and BERNET, E. H.

1945 Component indexes as a basis for stratification in sampling. *J. Amer. Stat. Assoc.* **40**: 330-341.

JESSEN, R. J.

1942 Statistical investigation of a sample survey for obtaining farm facts. *Res. Bull. Ia. agric. Exp. Sta.* **304** (104 pp).

JESSEN, R. J., and HOUSEMAN, E. E.

1944 Statistical investigation of farm sample surveys taken in Iowa, Florida and California. *Res. Bull. Ia. agric. Exp. Sta.* **329**.

JUNTA DE COLONIZAÇÃO INTERNA

1943 *Relatório dos trabalhos e Contas de Gerência* (ver ainda os Relatórios de 1944 e 1945), Lisboa.

KING, A. J., and JESSEN, R. J.

1945 The master sample of agriculture. *J. Amer. Stat. Assoc.* **40**: 38-56.

MAHALANOBIS, P. C.

1946a Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Stat. Soc.* **109** (4): 326-378

1946b Report on the Bihar crop survey: Rabi season, 1943-34. *Sankhyã* **7**: 29-106 (Culcutta).

MINISTÉRIO DA AGRICULTURA (Estação Agrária Central)

1934 A Cultura do Trigo na região do Alto Alentejo. Resultados de um inquérito económico e cultural. *Bol. Min. Agric.*, Lisboa **13** (A): 1-305.

MINISTRY OF AGRICULTURE AND FISHERIES

1944 Wireworms and food production. A wireworm survey of England and Wales. 1939-1942. *Bull. 128* (H. M. S. O., London).

MINISTRY OF AGRICULTURE AND FISHERIES

1946 *National Farm Survey of England and Wales.* (A summary report). London.

NEYMAN, J.

1934 On two different aspects of the representative method : the method of stratified sampling and the method of purposive selection. *J. roy. Stat. Soc.* **97** (4) : 558-625.

OLIVEIRA, A. J. DE

1948 Estudos de Estatística Agronómica. IV — Importância da amostragem na experimentação agrícola. *Agron. Lusitana* **10** (3) : 201-229.

UNITED NATIONS ORGANIZATION (Lake Success, U. S. A.).

1948a Recommendations Concerning the Preparation of Reports on Sampling Surveys (Ver : *Sankhyã* **9** : 392-398, 1949).

1948b *Program for the 1950 world Census of Agriculture.* Food and Agriculture Organization of the United Nations, Washington, D. C.

U. S. BUREAU of AGRICULTURAL ECONOMICS

1945 The BAE General-Purpose Sample of 101 Countries (ciclostilo).

U. S. D. of AGRICULTURE (Bureau of Agricultural Economics)

1949 The Agricultural Estimating and Reporting Services of the United States Departament of Agriculture. *Misc. Publ.* **703** (266 pp.).

VAZ, D. FARIA

1950 Estimativa das produções de vinho branco e vinho tinto nos anos de 1927 a 1936. *Sér. Estudos Inst. Nac. Estat.,* Lisboa **14**.

YATES, Fr., BOYD, D. A., and MATHISON, I.

1944 The manuring of farm crops ; some results of a survey of fertilizer practice in England. *Emp. J. exp. Agric.* **12** : 163-176.

YATES, Fr.

1949a *Sampling Methods for Censuses and Surveys.* Ch. Griffin & C. London.

1949b Agriculture, sampling and operational research. *Proc. Inst. Intern. Stat.*

ÍNDICE DO VOLUME XVI

FUNGI LUSITANIAE-VII — Maria Rosália de Sousa Dias et Emmanuele de Sousa da Camara 5

STUDY OF KARYOTYPES AND THEIR ALTERATIONS IN AROIDS — A. K. Sharma and N. K. Das 23

CHROMOSOMAL ABERRATIONS IN OATS, *AVENA SATIVA* L. — L. Costa Rodrigues 49

FUNGI LUSITANIAE-VIII — Maria Tereza Lucas et Emmanuele de Sousa da Camara 81

UMA SÉRIE POLIPLOIDE EM *LUZULA PURPUREA* — Duarte de Castro e Tristão Mello-Sampayo 111

GEDANKEN ZUR PHYLOGENIE DER MOOSE. *Eine entwicklungsgeschichtliche studie* — Dr. Heinrich Steinböck 115

RAÇAS FISIOLÓGICAS DE *PUCCINIA RUBIGO-VERA* F. SP. *TRITICI* (ERIKSS. & HENN.) CARL. ISOLADAS EM PORTUGAL — Alberto Palyart do Carmo e Freitas 151

FUNGI LUSITANIAE-IX — Aniceta Clotilde dos Santos et Emmanuele de Sousa da Camara 175

INSTABILIDADE CROMOSÓMICA EM *LUZULA PURPUREA* — Duarte de Castro 195

DINAMISMO CLIMÁTICO DA REGIÃO ALGARVIA. *No conceito de Lautensach & Freymann* — J. de Pina Manique e Albuquerque. 213

CORRECÇÃO DUM SOLO HALOMÓRFICO. *Estudo laboratorial comparativo entre o Krilium, flotal e gesso* — L. Rodrigues Balbino. 231

ESTUDOS DE ESTATÍSTICA AGRONÓMICA-V. *Os modernos inquéritos estatísticos por amostragem no estudo dos problemas agrários nacionais* — Augusto José de Oliveira 283

PUBLICAÇÃO DA
DIRECÇÃO GERAL DOS SERVIÇOS AGRÍCOLAS
Repartição de Estudos, Informação e Propaganda

PUBLICAÇÃO DA
DIRECÇÃO GERAL DOS SERVIÇOS AGRÍCOLAS
Repartição de Estudos, Informação e Propaganda

SUMARIO

Oliveira, Augusto José de — ESTUDOS DE ESTATÍSTICA AGRONÓMICA - V. OS MODERNOS INQUÉRITOS ESTATÍSTICOS POR AMOSTRAGEM NO ESTUDO DOS PROBLEMAS AGRÁRIOS NACIONAIS 283-346

TODA A CORRESPONDÊNCIA E PERMUTA REFERENTE À AGRONOMIA LUSITANA DEVE SER DIRIGIDA A:

POUR L'ÉCHANGE OU POUR LA CORRESPONDENCE CONCERNANT L'AGRONOMIA LUSITANA ON DOIT S'ADDRESSER À:

LETTERS AND EXCHANGE CONCERNING AGRONOMIA LUSITANA TO BE ADDRESSED TO:

ESTAÇÃO AGRONÓMICA NACIONAL
SACAVÉM
PORTUGAL